Decolonizing the Data Science Community through Meaningful Inclusion of Underrepresented Voices

Tian Qing Yen
CDI Master's Thesis
October 31, 2023

TU Delft

Decolonizing the Data Science Community through Meaningful Inclusion of Underrepresented Voices

by Tian Qing Yen

5471079

Master's Thesis in Communication Design for Innovation
Delft University of Technology

Supervisors: Dr. É. Kalmàr
Drs. C.Wehrmann
Dr. T. Verma

31 October 2023

Acknowledgements

I could write an entire thesis about all the people who have been an integral part of this process. The words that I have written are not mine alone; they are ideas sprouted from countless conversation with others. This was a community effort from beginning to end, and I am infinitely grateful for everyone who has supported me along the way, for all the people who put their trust in me, even when I didn't trust myself.

Thank you to all of the wonderful people with whom I talked for this research. To the professors, postdocs, PhDs and students who took time out of their busy schedules to answer my abstract questions: your stories have impacted me far beyond what is written here and I hope that I am able to represent you in a way that is true to your experiences. To the two lovely people who provided feedback for my framework: thank you for trusting me with your stories. Your vulnerability challenged me to look critically at my own purpose and the impact of my work. Your experiences motivated me to try and create something meaningful – if only because you deserve a system far better than the one we have now.

Thank you to my friends and colleagues who planted new seeds in my head with your thoughtful questions. To CUSP: you welcomed me in to your group with curiosity and joy. It was an incredible experience to be surrounded by people who all believe so strongly in building a more just and equitable society. It showed in the kinds of questions you asked and how you challenged each other, myself included. This group is already enacting so many aspects of inclusion that I write about in this thesis. To the legendary CDI students in the afstudeerhok: from impromptu conversations and coffee breaks to feedback sessions to inspiring memes on the wall, you all provided the company, compliments and snacks I needed to keep going. Thank you for ranting with me, crying with me, and laughing with me. Your camaraderie made this process (at least occasionally) something incredibly joyful.

Thank you also to my classmates, friends, and mentors who helped me branch out and connect with others. To Tanya from the Open Science Community: thank you for connecting me with my supervisor and for continuing to support me throughout the process. Your enthusiasm and insights were like sunshine when I felt most stuck. To Charlotte from the AI Initiative: you dug into my project without much context and provided me invaluable feedback about the community. I hope that this project can give back to you all as much as you have provided me.

Thank you to Roberta, my lifelong friend and mentor for guiding me both in life and in this project. You were there to support my big steps, from deciding to move to the Netherlands, to choosing my research focus for this project, to wrapping everything together in my discussion. I am so grateful you were part of this journey.

To my supervisors: Éva, Caroline and Trivik, thank you for providing the solid ground for me to do this work – especially when I felt the turmoil of writing a thesis most profoundly. You created space for me to follow my intuition and lead you along this journey as well. Trivik, you presented me with an idea and encouraged me to run with it. From the very beginning, you had so much faith in my experience and expertise. Thank you for encouraging me to share my ideas and welcoming me into CUSP. Éva, during our very first meeting, you asked me what kind of support I like. I told you that most of the time, I just need a bit of positive encouragement. You provided a trellis on which I could grow, nudging, encouraging and inspiring me every step of the way. Thank you for your kindness and support. Caroline, you see me as a whole human and I am so grateful for that perspective. Thank you for always reminding me to see my own

strengths. You all embody the kinds of change that I write about in this thesis. I think the university can learn just as much from the way that you three supervise, care for, and nurture students like myself.

Thank you to my frisbee team, Force Elektro: you always see the best in me during the moments when I am most tired, stressed, and frustrated. Thank you for providing the encouragement, connections, and community I needed to finish this thesis. The endless supply of hugs at practices kept me going strong until the end.

Thank you also to my people who continue to support me from afar – you make me feel loved and connected at all hours of the day. My Beijing girls, always willing to listen to good rant. My Cozy Lounge co-working buddies, who provided me company on days I worked too long, asked such good questions I had to write them down, and made sure to periodically check, "ARE YOU ALIVE!?"

Thank you to my house, the Rolandians (past and present): I feel so lucky to have spent the last two years with you all. From house dinners to MarioKart to walks around the neighborhood, I am grateful for all the little things that make our home feel like home.

To Sandy and Nick: thank you for reminding me that life is much bigger than just one project. You helped me to see the horizon and imagine what might come next and to stay grounded and take each day as it came.

Finally, to my Baba, you have always been my rock. Thank you for showing me love and support in my lowest moments, and always encouraging me to follow my dreams.

Contents

1 | Introduction

Perspective Matters

About a month into my thesis, I met up with a friend for coffee. This friend told me about a course they took on Sustainable Engineering. The curriculum for this course drew inspiration from a transdisciplinary workshop[1] that incorporates critical discussions about the social and ecological responsibility of engineers. During this course, the students participated in an activity where they were asked to assume the perspective of different stakeholders in a conversation about a lithium mining project in Chile. One student in each group was assigned the role of an impartial decision-maker whose purpose was to listen to these perspectives and provide recommendations for the project. By the end of the activity, every group arrived at the same decision: we need more regulations, better compensation for the affected locals, and more protection to minimize the environmental harm.

This raises a question: in a situation with numerous potential solutions, how did every group arrive at the same one? My friend, upon noticing this, challenged the class to acknowledge their bias towards solutions that are relevant in a European value system and regulatory context. Having grown up in a non-European context, my friend recognized how our culture, values, and surrounding political climate can impact what we consider to be a fair and reasonable decision. A politician in Chile with different values would evaluate the project and stakeholder perspectives differently from the students in this class. The purpose of the activity was to consider the context in which we engineer solutions, and yet no one considered the context of the most important role: the decision-maker. Instead, the students in this role took on the perspective they were most familiar with; they took on the dominant perspective in the room. Decision-makers always operate in a context; yet, by framing this role as "impartial" we are implicitly teaching these students that a Eurocentric perspective is an objective perspective. At best, this undermines the validity of non-European perspectives by positioning anything else "subjective." At worst, this framing excludes these other worldviews from entering the classroom entirely.

This story demonstrates how easily perspectives can be erased from academic spaces. Despite attempting to involve a diversity of global challenges and consider alternative viewpoints, an intention to be inclusive can have the opposite effect of creating a more exclusive and alienating environment.

1.1 Project Background

We live in a society where our decisions are increasingly based on research, data, and modeling. The role of research institutions – including Delft University of Technology (TU Delft) from which I am writing – is to both educate future engineers and to produce research outputs that support a data-driven society. However, research is an ongoing way of regulating and realizing the underlying codes of imperialism and colonialism (Smith, 1999). Universities are especially complicit in this perpetuation of asymmetrical power relationships by occupying a

[1] More details about the original workshop offered by Berlin Technical University can be found at http://www.blue-engineering.org/wiki/English [Accessed: 12 Sept 2023]

role of "well-intentioned post-colonial saviors" (Anthony-Stevens & Matsaw Jr, 2020, p. 596). This positioning is embedded in TU Delft's strategic vision and mission to **educate socially responsible engineers who can contribute to solving global problems in collaboration with societal actors**.[2] This vision is demonstrated in courses such as the one I described in my friend's story; it also embodies how universities maintain forms of colonialism and oppression through 1) the framing and construction of knowledge and 2) the misrepresentation and alienation of non-dominant groups in research. A consequence of such ongoing inequity is that it also undermines our ability to collaborate with societal actors. If we truly have an intention to make "our research and innovations more inclusive"[3] then we must also critically consider our role in upholding colonial systems. Starting with an examination of how colonialism shapes knowledge, representation, and collaboration, I attempt through this thesis to redefine meaningful inclusion within one part of TU Delft – the data science community – through a lens of decolonization. TU Delft is soon due for a new strategic vision, and I hope that this perspective challenges us to reconsider the disconnects between our ideals and the realities we reinforce.

1.1.1 Recontextualizing a Decolonial Perspective

Before I dive into literature and frameworks on decolonization, I want to recontextualize the knowledge from which I am drawing. Decolonization theory is rooted in the work of numerous indigenous scholars and indigenous methodologies, including the formative work of Linda Tuhiwai Smith in *Decolonizing Methodologies: Research and Indigenous Peoples* (1999). However, as Evans et al. (2014) caution, only understanding indigenous research in relation to what it is not (i.e. colonial/European) can be a potentially recolonizing act. Given that my own research scope does not actively uplift the experiences of indigenous peoples, nor prioritize research that benefits those communities (Beeman-Cadwallader et al., 2012; Datta, 2018; Evans et al., 2014; Smith, 1999), I doubt whether to consider my work decolonizing at all.

This is a question that is not unique to indigenous scholarship, but to many forms of oppositional scholarship that challenge dominant ways of knowing. Patricia Hill Collins, a prominent scholar in Black feminist thought, writes about how to remain oppositional even as Black feminist scholarship is given greater visibility and ostensibly included within social institutions. Collins (2016) emphasizes that small decisions about our practice can matter just as much as broad social movements. Throughout this thesis, I return to this question to reflect on my decisions: *what does it mean to remain oppositional?* As I discuss in Chapter 3, this thesis is also a story about decolonizing my own research process.

For me, this starts with incorporating forms of knowledge that are not generated through the academic peer-review system. I include books, blogs, essays, podcasts, and art as sources, especially those that are created by groups who have been historically excluded from academic spaces. In doing so, I also acknowledge that the experiences of people – scholars, artists and everyday people alike – are representative of their understanding of the world; I give visibility to their perspectives as well.

Part of visibility is also recognizing and celebrating the heterogeneity that exists within art and scholarship on intersectionality, social justice, Black feminist and transnational feminist work, indigenous research and ontologies, and decolonizing methodologies. The communities,

[2]Paraphrased from the TU Delft's vision and mission statements, published in TU Delft Strategic Priorities 2022-2024 (p. 12) [Accessed: 3 Oct 2023]

[3]TU Delft Strategic Priorities 2022-2024 (p. 44) [Accessed: 3 Oct 2023]

knowledge systems, traditions, and perspectives I refer to encompass a vast diversity and mean different things around the world. Evans et al. (2014) explain that indigenous ontologies (i.e., the knowledge derived from ways of knowing, being and doing) are both metaphorical and symbolic and thus must be understood within the cultural, geographical, and linguistic contexts. Especially when I draw on specific practices, concepts, and ideas from indigenous scholars, I try to acknowledge this heterogeneity by including the Nations or Peoples with whom these scholars are working or to which these scholars belong. This is an oppositional practice to counteract the erasure of indigenous peoples through assimilation and conglomeration.

1.1.2 Deconstructing University Power through Decolonization

A Dominant Perspective of Knowledge

Universities act as gatekeepers of knowledge, positioning themselves as experts within society. Knowledge and power are deeply intertwined in colonial systems (Evans et al., 2014). Deconstructing the role of universities in perpetuating such unequal systems starts with understanding the power universities possess to construct, frame, validate, and thereby perpetuate certain ways of knowing.

Throughout my education, I have been taught that the scientific method is the primary way through which we can understand the world. However, the belief in an objectively knowable world is a fundamentally European perspective. Ndlovu-Gatsheni describes how Cartesian Duality, the philosophical underpinning of European Enlightenment thinking, led to a conception of knowledge that is "objective, unsituated, and universally truthful" (2019, p. 484). Embedded in the phrase *I think, therefore I am* is a separation of body and mind, of subject and object, that was used to justify an era of "discovery" and "othering" through colonization (Ndlovu-Gatsheni, 2019; Smith, 1999). This frame of mind allowed those in power to distance themselves from the environments and peoples they subjugated. It created a narrative in which indigenous peoples were cast as inferior and the land was viewed as a resource to be taken and used. As Smith (1999) points out, the concept of distance is also important in research because it implies neutrality and objectivity on behalf of the researcher. It conflates something that is measurable (distance) with something that is not (objectivity). This in turn perpetuates a myth that distancing ourselves from the subject of our research is a way to be more objective.

The scientific method that I learned, that is often employed in the research of natural sciences, is rooted in rationality. Modern rationality is dependent on dissociating from the body; the body becomes an object separate from subject, spirit or reason (Quijano, 2000). This separation of mind from both the physical and spiritual worlds leads us to disregard knowledge that is rooted in spirituality or experience. Biologist and author Robin Wall Kimmerer of the Potawatomi Nation writes, "...science is rigorous in separating the observer from the observed, and the observed from the observer. [Asking] Why two flowers are beautiful together would violate the division necessary for objectivity" (2013, p. 42). This quote from her book *Braiding Sweetgrass* highlights a tension between her cultural knowledge and what she was taught qualifies as "objective" scientific knowledge in her education. The subjective concept of beauty is not considered relevant to biology research; it belongs to other disciplines such as philosophy or art. Jama (2023) experienced a similar tension working in software development and employing a form of rationality which ignored contextual, relational or emotional considerations. They described the contradiction between doing work that is rational and doing work that was meaningful as a profound spiritual violence. In both of these examples, what might be considered human, subjective considerations were separated from technological spaces or scientific

research. This separation of disciplines – thinking in disciplines to begin with – is a way through which academia can decontextualize knowledge.

A disciplinary way of thinking is also a consequence of imperialism, which disconnected the histories of indigenous peoples from the landscape, language, cultural knowledge and interactions they had with the world (Smith, 1999). Even today, the structure of disciplines and the management of information within academia can have real consequences on how the histories and perspectives of indigenous groups are remembered. Nathan et al. (2017) recounts an example[4] from the University of British Columbia (UBC): there is no section in UBC's library catalogue for Musqueam materials, despite that the university is built on Musqueam territory, Musqueam people teach and learn at the university, and Elders from this community consult and sit on advisory boards. By not including "Musqueam" as a category in their information management system, the university is continuing to separate and decontextualize the interconnected knowledge belonging to this Nation. It also impacts how Musqueam people are seen as their own distinct Nation possessing a unique culture.

Decontextualization is a pervasive and ongoing part of Western society. It allows for the continual dehumanization of subordinate groups by stripping them of the ability to shape their own story, or by constructing knowledge around their status as subordinate. Collins (2016), writing as scholar in the US, describes how depictions of scholarship, media, and other main-stream knowledge are intertwined with the subordination of Black women. Our cultural knowledge in society is dependent on the continual undermining of Black women and their experiences. Aníbal Quijano (2000) refers to this as the "coloniality of power": the cultural, political and economic processes through which oppression of a subordinate group is maintained and Eurocentric ways of knowing and being are reproduced as superior. The story I began with is also an example of the coloniality of power in action; it is an example of upholding the impartiality of a Eurocentric perspective. While Quijano wrote about manifestations of a colonial legacy in Latin America at the turn of this century, certainly the coloniality of power persists in our increasingly globalized world. Within a US context, hooks (2003) writes about a "dominator culture" characterized by cynicism, competition, and the belief that domination is the foundation of all human relations. As with the coloniality of power, dominator culture reinforces a dualism in which you are either dominant or subordinate, validating systems in society which operate by oppressing others.

These concepts are visible in the underlying culture of "Western research," which seeks to create systems of classification and representation (Smith, 1999). Datta (2018) describes this as an "invisible power" embedded in the way researchers define, essentialize, label, and thereby alienate others. This is one way that research attempts to frame knowledge as objective truth. Yet, as Māori scholar Smith explains, "we are not the final arbiters of what really counts as truth" (1999, p. 34). Instead, "truth" is determined through peer-review processes and academic discourses that deny multiplicity in ways of knowing and restrict who can participate. The peer review process itself is not unbiased; it can replicate the same power imbalances seen in other parts of society with potentially dangerous consequences. Sultana (2018) writes about an instance from 2017 where the editor-in-chief of the journal *Third World Quarterly* published an article arguing for the recolonization of Third World countries, in spite of the reviewers who rejected the article for being both academically and ethically ungrounded. This editor-in-chief was incentivized to publish an inflammatory "opinion" article to stir up debate because they also

[4]Originally quoted from: Ann Mary Doyle. 2013. Naming, claiming, and (re)creating: Indigenous knowledge organization at the cultural interface. Dissertation. University of British Columbia. https://dx.doi.org/10.14288/1.0073667

earned money directly from the journal. This one person overrode the purportedly "unbiased" academic peer-review system. Publishing this article which expounded benefits of colonization while ignoring its destructive history had the consequence of emboldening white supremacists by legitimizing their racist ideologies. This is a modern-day example of attempts from within academia to re-write history and further entrench power asymmetries in society.

There are also subtler ways in which the myth of objectivity in research can result in harms to society. Claiming neutrality limits our "capacity to uncover (analytically and conceptually) the ways in which our research is gendered, age-biased, color-blind, or Global North-centered" (Franklin et al., 2023, p. 3). When we talk about research for "the public," the public typically refers to dominant groups in society (Smith, 1999). This has had real consequences in how we understand the world. For example, medical devices tested on white-skinned people often have a much higher inaccuracy for people with non-white skin, effectively limiting their ability to access this potentially life-saving technology (Parthasarathy & Stilgoe, 2023).[5] This is just one example of how biased research can perpetuate structural inequities in society. As Nathan et al. (2017) criticize, we cannot assume that technological interventions contribute to a universally shared common good, or that researchers have the expertise to determine what is best for all people.

For a project to be decolonizing, it must recognize the implicit processes through which research produces and legitimizes knowledge (Beeman-Cadwallader et al., 2012). Datta (2018) defines decolonizing research as both 1) understanding the history of colonialism and how it has shaped our educational systems, research practices, and relationship to science, data, etc. and 2) recovering and actively counteracting these impacts by uplifting the voices, perspectives, and experiences of indigenous peoples, thereby solidifying both their place and identity. Digging into the interconnected history of research, knowledge and colonization, we can begin to question the ongoing role of the coloniality of power and dominator culture in shaping universities. To understand this second aspect of decolonizing research – solidifying place and identity – we must also look at the ways universities and research create a culture of alienation and erasure through conformity.

Inclusion in a Culture of Conformity

Earlier this year, an associate professor in Aerospace at TU Delft resigned due to what news platform *TU Delta* called "macho culture." She explained her decision as, "the end of a very big story in which my expertise in a particular subject was never recognised. Decisions affecting me were taken without consulting me."[6] This is an example from our own institution in which the people in power keep the same people in power. This is an example of dominator culture in action, a culture hooks explains "has tried to keep us all afraid, to make us choose safety instead of risk, sameness instead of diversity" (2003, p. 197). Without addressing this ingrained inclination towards sameness, conversations about diversity and inclusion remain just that – conversations.

[5] A specific example of this is the pulse-oximeter, an affordable technology that measures blood-oxygen levels through the skin. During the Covid-19 pandemic, it was seen in the U.S. as a potentially life-saving technology because it allowed people at home to identify when they might need additional medical attention. Scholar Amy Moran-Thomas talks about the biases embedded in the development of this technology and its implications on the podcast The Recieved Wisdom, Episode 32: The Politics of Expertise and Retelling the Story of Racism in the Pulse Oximeter ft. Amy Moran-Thomas (19 March 2023)

[6] Quote taken from a TU Delta article: "Associate professor Daphne Stam stepped down over 'macho culture' at Aerospace Engineering" published August 29, 2023 [Accessed: 5 Sept 2023]

It is becoming a commonplace practice for institutions to have some kind of diversity and inclusion policy. Diversity is one of TU Delft's core values. Yet, even if we welcome the idea of diversity, we do not necessarily give space for diverse perspectives to exist in academia. Knowledges coming from Indigenous Peoples, minorities, and cultures of the Global South in particular are excluded from both STEM and higher education discourses (Anthony-Stevens & Matsaw Jr, 2020; Chan et al., 2020). This is evident in the resignation of the aerospace professor – her knowledge and input were not recognized by those already in power. Existing power relationships in society and status in research influence the way knowledge is constructed and disseminated (Chin et al., 2022). More often than not, this provides privileged groups in society – the white men of "macho culture" – the ability to gatekeep which perspectives are validated, prioritized and thereby maintained in academic spaces. Culturally constructed systems that have centripetal movement place the burden on new participants to reimagine and adapt their practices to the existing system in order to move towards central participation (Fish et al., 2022). In other words, to be in a position of power within a centrally operated system, you must already fit into that system.

Universities demand a conformity that is fundamentally counter to efforts at inclusion. Writing from their experience working as a Black woman in the US tech industry, Jama asks rhetorically, "But what do inclusivity and diversity matter when you have to mutilate yourself in order to fit these spaces?" (2023). Kirkness and Barnhardt (2001) share experiences from First Nations students attending universities in the US and Canada who felt this pressure firsthand. One student explained, "I would like to tell them (at the university) that education shouldn't try and make me into something I'm not" (2001, p. 5). The authors argue that universities demand individuals become integrated into the existing academic and social culture and aligned towards reproducing similar goals. Datta (2018) describes their own PhD training as focused on epistemological position and the way researchers see their function as researchers. In teaching about what it means to be a researcher rather than how to be a researcher, we leave less space for people to bring their own identity into their work or pursue what is meaningful and relevant to them. Chin et al. (2022) explain that the onto-epistemological assumptions a researcher makes influence every aspect of research and design. When universities teach epistemological position, they enforce a research culture and research agenda based one perspective. A challenge for universities truly seeking to be diverse is to disrupt our orientation towards one perspective by including other ways of knowing in our education and research.

However, attempts to include other ways of knowing in academic discourse are insufficient on their own. Anthony-Stevens and Matsaw Jr (2020) reflect on their experience incorporating Indigenous knowledge systems from peoples in the Northwest of the US into an interdisciplinary graduate course. They found that students, despite finding Indigenous ways of knowing valuable to contextualize their studies, avoided engaging with the power and coloniality rooted in their own research. Similarly, hooks (2003) observed that educators supported including diverse ontologies in their classrooms only when these ontologies were taught as inferior to dominator culture. Collins argues that in "an era characterized by politics of inclusion," it is even more important that we remain oppositional to conventions that would undermine alternative ways of knowing (2016, p. 135). This oppositional nature, actively challenging the dominant narrative, can make inclusion a decolonizing practice. Falcón (2016) defines decolonization as destabilizing the dynamics that privilege dominant groups or ways of knowing. This requires not only including non-dominant ways of knowing in academia, but also centering and validating them.

Inclusion through Reflection and Representation

Focusing on inclusion is important for a university that also strives to collaborate inter- and transdisciplinarily. Evans et al. (2014) found in their research with a Métis community that effective inclusion increased community capacity and adaptability – i.e. the community's willingness to accept interventions coming from outside their community. For Evans et al., effective inclusion involved promoting cultural continuity and sustainability through the use of community experts and resources. By drawing on the knowledge already within the community and respecting cultural practices, these researchers were able to validate the perspectives of the community and earn trust. Meaningful involvement and ethical relationships with collaborators can be a way to purposefully represent a community through your research (Beeman-Cadwallader et al., 2012; Datta, 2018)

Representation is a concept that shows up in decolonizing methodologies and social justice frameworks. Fish et al. (2022) explain that representation is the difference between having snippets of your culture added to your education, and having your cultural knowledge seen as valid and valuable. Being represented gives an impression of truth when your reality is reflected back at you (Smith, 1999). Acknowledging and valuing different perspectives is a form of justice and a way to include groups in a more equitable way (Blue et al., 2019). Nancy Fraser, a feminist and theorist, considers representation to be a prerequisite for social justice. In addition to being recognized as valid by others, representation is also the capability to make yourself heard and to contribute in your own voice (Rosa, 2017). Representation includes a sense of agency to shape your own story and to make decisions in your life (Musara et al., 2021). Perhaps an equally important aspect of inclusion is our ability to reflect our own reality back at ourselves, to validate our own position and truth.

A decolonizing form of inclusion is one that allows us to bring into spaces our full human complexity. Reflection is a process through which we can create space to integrate divergent ways of knowing by seeing knowing as an ongoing continuum (Anthony-Stevens & Matsaw Jr, 2020). Reflexivity allows us to identify other's perspectives and shape how we represent ourselves. "Being aware of complexity is an act of self-preservation" (Jama, 2023). In research, holding space for the intersectionality of identities also encourages accountability because we learn how to appreciate and thrive in complexity (Boyd, 2021).

Reflexivity is one way we can create space for the complexity and multiplicity of truths humans carry. By creating that space, we validate and represent our unique perspectives. Representation in turn facilitates more meaningful and effective inclusion that allows for collaboration and accountability in research. This process ties back to decolonization, requiring we first understand how colonization shapes our educational system and construction of knowledge.

1.1.3 Colonial Patterns in Data Science

We live in increasingly digitized societies where our economies, technologies, and politics are shaped by data science. Yet data science, like academia, is deeply intertwined with colonial history and power asymmetries. Digital society reflects the physical world, inheriting historical patterns of inequality (Bon et al., 2022). Detangling data science and colonization can challenge these patterns of inequity, allowing us to "build a more human-centered, participatory, and democratic digital society – inclusive also for the most vulnerable communities" (Bon et al., 2022, p. 67).

I chose to focus on decolonizing the data science community within TU Delft for a few reasons. Data science and digital technologies are an undeniable part of our future; the start of this academic year 2023-24 was aptly named "Entering the Age of AI."[7] This field is growing at an incredible speed, which makes it even more important to critically consider its role in excluding, exploiting, or otherwise harming certain groups of people. As with academia, the coloniality of power and dominator culture appear in data science and promote a Eurocentric and Western way of knowing. A way this occurs is by making invisible the harms and oppression on which digital technologies are often built. The invisibility and erasure of such harms in turn impacts our ability to create technologies that benefit groups that are less privileged in society. Data science is inherently multidisciplinary; if our university intends to create inclusive technologies, it must start with an inclusive research community.

Dominance through Data Colonialism

Dominator culture and the coloniality of power operate by reinforcing the superiority of one truth. This culture appears in data-driven fields as well. The rise of big data was accompanied by the myth that these data represent an "austere, predictive truth" that can enhance insights and generate cost-effective solutions (Thatcher et al., 2016, p. 992). Data are falsely thought to be objective. However, Franklin et al. (2023) challenge us to consider who is missing in spatial data sets, as those missing are often already underrepresented. Research methods based on data, conceptual frameworks, and quantitative methods all "have the power to exclude" and risk reinforcing systemic privilege (Franklin et al., 2023, p. 5). Thatcher et al. (2016) use the metaphor of data colonialism to emphasize the ways in which data collection and processing lead to power asymmetries. In her book *Race after Technology*, Benjamin (2019) describes numerous examples of data colonialism; she demonstrates a pattern in which companies that control and ultimately profit from data do so with disregard to the rights and benefits of the people from whom data are extracted. The generators of data are denied access to the value produced from it through accumulation, privatization, and aggregation (Thatcher et al., 2016). Data is a form of knowledge. How we define and extract data reproduces the same decontextualization apparent in other fields of research. Universities, as keepers and producers of data, must also ask ourselves whether we are reinforcing unequal distribution and access to these data.

Data colonialism also operates by centering the control and regulation of digital technologies in the Global North. Software and digital platforms are overwhelmingly based in the US and Global North, allowing these regions to exert control through forcing Indigenous, rural and Global South populations to adapt to their practices to gain access (Meital & Jason, 2022). The widespread nature of digital technologies also leads to a system in which individual users may be harmed or isolated by refusing to conform to the system (Bon et al., 2022). Data colonialism is an alienating process because it forces conformity to an existing system and hides the way individuals or groups are harmed by the system.

Invisibility in Data Science

Data science suffers a challenge of invisibility. Western research practices are effective in separating subject from researcher, and that is even more evident in data-driven research. Especially when working with characteristics such as demographics, "there are social choices embedded in the socio-demographic characteristics we choose to measure; simply because we do not measure

[7] From TU Delft's Opening Academic Year 2023-2024 website [Accessed: 14 Sept 2023]

it, does not mean it does not exist" (Franklin et al., 2023, p. 4). The choice to not include a particular group, however, limits our ability to design technologies that benefit those groups. For this reason, Bon et al. (2022) emphasize that any efforts to decolonize digital technologies require collaboratively involving the voices and perspectives of communities in the Global South, or others who are excluded from digitalization.

Invisibility also occurs when we fail to acknowledge the ways digital technologies privilege dominant groups. A lack of representation in the construction and function of these technologies produces ongoing biases that exclude non-dominant perspectives. For example, Gasparotto (2016) explains that search engines deprioritize the visibility of materials created by and for indigenous peoples, a bias that resulted from indigenous communities being physically displaced in the Americas which made it challenging to acquire internet infrastructure early on. Meital and Jason (2022) describe how the dominance of English in digital landscapes limits the accessibility of peoples in Sub-Saharan Africa to engage in these spaces, often because their linguistic diversity is not acknowledged and supported in these platforms. There is privilege and power that comes with being able to utilize materials from and operate in digital spaces.

Not only do digital technologies privilege dominant groups, they also often do so at the expense of others. An enormous amount of resources go into producing and maintaining technologies that we take for granted – low-paid workers who train search algorithms (ghost workers), the energy required to run large data centers and servers that host these technologies. The danger of modern technologies is that they "cultivate a form of unawareness that divorces us from the value of what we interact with and the people who spend an excruciating amount of effort creating said value" (Jama, 2023).

Data science methods themselves also require that we simplify the reality, flattening and distorting real experiences to an oversimplified model (Boyd, 2021). We decontextualize data in the process of attempting to quantify activities (Thatcher et al., 2016). To counteract this decontextualization of reality, we need to acknowledge, to unforget, those whose quality of life suffers for our ease and convenience (Jama, 2023). We need to make the invisible visible.

Universities have an opportunity to shift away from an alienating research paradigm towards a form of data science that is inclusive. Data science is often seen as a democratizing science, having the potential to both reinforce existing biases and social stratifications and to counteract them (Benjamin, 2019). Yet, most Computer Science and Artificial Intelligence programs do not focus on collaborative development of community-centered technologies or on developing technologies in resource-constrained environments (Bon et al., 2022). Similarly, technologists are not trained to work cross-culturally and end up in workgroups with people who share the same experiences, values, and beliefs. Just as hooks (2003) observed educators choosing sameness, Turner Lee (2018) observed that people working in machine learning and AI in the US also exhibited an "inattentional blindness" to self-select similar perspectives. Jama (2023) aptly recognizes that solutions to the challenges of the technological revolution are not going to be found in the technology itself, but rather in the relationships we have with each other and with these technologies.

It is ambitious to imagine deconstructing the coloniality of power within data science and digital technologies as a whole. But starting with data science research at TU Delft – looking critically at how inclusive our own relationships are – might be a good place to start.

1.2 Project Purpose

Both TU Delft and the field of data science are complicit in perpetuating power asymmetries that exist in society. However, this also means academic institutions have an opportunity to challenge the dominant narrative. This thesis offers a perspective on inclusion that can support efforts to decolonize the data science community at TU Delft. The overall purpose of my research is to understand meaningful inclusion in this community as a starting point from which to begin dismantling the legacies of colonialism in academic spaces.

1.2.1 Research Questions

With this goal in mind, my overall research question is:

To what extent are the perspectives of people from backgrounds that have historically been (or continue to be) excluded from academic spaces meaningfully included within the data science community at TU Delft?

To answer this question, I look at four dimensions:

1. **How can position within the data science community at TU Delft be described through relationships to people and processes?**

   Through this question, I am attempting to establish the boundaries of the data science community and where people are situated within those bounds. I look at positionality through relationships because cultivating relationships is an essential part of decolonizing research (Beeman-Cadwallader et al., 2012; Datta, 2018). I consider relationships to people, to processes and to institutions, all mentioned as forms of interaction by Rosa (2017).

2. **To what extent does position in the community impact the ways people form relationships, access resources, or participate in processes?**

   This is a question about where momentum is occurring in the community. Communities are constantly shifting. With this question, I try to connect how the community is structured to what is happening in the community.

3. **To what extent do people from various identities feel as though they can express their perspectives in the community?**

   This question returns to the aspect of representation that is concerned with our own agency, voice, and telling our story. Storytelling is a decolonial act that upsets the decontextualization of subjective human experiences (Anthony-Stevens & Matsaw Jr, 2020). By looking at representation also through our ability to tell stories, I am evaluating potentially decolonizing actions the data science community may already be taking.

4. **Finally, how can a decolonial perspective along with feedback from the data science community itself be combined to meaningfully include the perspectives of those positioned at the edges of this community?**

   The questions about position lend insight into who might be located at the edges of the TU Delft data science community. Asking about perspectives is a way to assess whether diverse worldviews are represented. With representation, however, often the question of

> "'who will speak' is less pertinent than the question 'who will listen?'" (Gayatri Spivak as qtd in Smith, 1999, p. 71) What is left to ask is: how might we create space for these perspectives to be shared at a community level?

In asking these questions, I am starting from a set of assumptions. First is the assumption that TU Delft and the data science community within TU Delft value diversity. On paper, it certainly seems so. Second is the assumption that this community contains diverse perspectives at all. I consider the extent to which people feel capable of expressing their perspective without considering to what extent there even are differing perspectives. The latter question asks for a more thorough consideration of who belongs to this community than what I could accomplish in this thesis. So instead, I assume that there are some perspectives that are over-represented in the values, processes, and structure of the community. Finally, although I lay out an argument in the introduction as to how including a diversity of perspectives can be a decolonizing act, I do not assess in this research whether inclusion is in fact decolonizing. I start from the belief that the community will benefit from meaningful inclusion, and believe that is enough justification to do this work.

1.2.2 Project Overview

The remaining six chapters of this thesis describe my journey to try and answer the research questions I have presented. In Chapter 2 I start by defining meaningful inclusion. I discuss how this definition shaped both my sub-questions and my approach to answering those questions. This framework combines concepts from decolonizing, indigenous, and feminist literature, along with feedback from two students at TU Delft who have experienced exclusion personally.

Chapter 3 details my overall research process. My process began with a literature search and co-creation sessions to get feedback on my framework (Section 3.1). This framework informed the kinds of questions I asked in interviews with people who are part of the data science community at TU Delft. I describe in Section 3.2 how I recruited for and conducted these interviews. Finally, I give an overview of the design steps that I took to incorporate these findings into an actionable tool for the community (Section 3.3). Throughout the chapter, I reflect on limitations of my process and highlight major decisions I made along the way.

In Chapter 4 Chapter 4, I present detailed findings from my interviews. The results are split into seven thematic sections, each followed with a brief discussion to connect back to my framework and other results. Section 4.2 attempts to answer my first sub question about describing positionality in the community. Section 4.5 and Section 4.6 explore an answer to the second sub question by presenting participants' thoughts on relationships, resources, processes, and transformations. My third sub question is addressed in Section 4.3 and Section 4.4. The last section (Section 4.7) considers community movement and primarily supports sub question 4 and the development of a design.

Chapter 5 is dedicated entirely to answering sub question 4. I introduce an activity I designed for the community. I explain how different results informed and supported this design. In Chapter 6, I make an argument for how this community can approach inclusion more broadly and reflect on the impact of my project. Finally, in Chapter 7 I conclude by returning to my main research question.

2 | Theoretical Framework

In the introduction, I hint at an interpretation of inclusion that is based on a decolonial perspective. In this section, I create a framework through which to understand meaningful inclusion in academic spaces. What is necessary to make people feel like they are heard, understood, and can participate fully in a community? How can inclusion be conceptualized in a way that places emphasis on empowering and centering the voices of underrepresented individuals?

I explore a version of inclusion by combining ideas from three works written by indigenous and feminist scholars. First is a framework for social justice from U.S. theorist Nancy Fraser (2008), as described by Blue et al. (2019), Musara et al. (2021) and Rosa (2017). The second work is from Falcón (2016), who lays out three transnational feminist principles that can support decolonizing methodologies for scholars researching across geopolitical borders. The third work from Kirkness and Barnhardt (2001) provides a vision for higher education in the U.S. and Canada that is inclusive to First Nations people. These works provide a foundation to guide my methodology and approach to answering my research question. I also bring into this framework ideas from two co-creation sessions with students to ground my theoretical understanding of inclusion with real experiences of people at TU Delft (Section 3.1.2).

2.1 Inclusion as Participation

Nancy Fraser lays out a framework for justice based on the parity of participation. Fraser argues that social justice depends on having social arrangements in society which allow for all people to participate equally as peers. These social arrangements ensure both the right to participate and the capability and opportunity to do so (Rosa, 2017). Musara et al. (2021) reflect on the relevance of this framework for understanding inclusion in education systems in South Africa. They explain that inclusion is not only about making everyone welcome in a space, but creating a space that gives voice to all involved by questioning patterns of power and disempowerment. By welcoming people into a space, we accept the normative value that people should be able to participate; questioning patterns of power and disempowerment then prompts us to consider whether people actually have the capability or opportunity to do so.

Fraser makes this framework actionable through three interconnected concepts that each address a different dimension of society: redistribution, recognition, and representation (Figure 2.1). Blue et al. (2019) use these concepts to guide inclusive participation when engaging stakeholders in planning for climate change. *Redistribution* highlights how the unequal allocation of resources and opportunities in society creates barriers for participation. *Recognition* addresses the way social status can produce and sustain inequalities in society that might also hinder engagement. *Representation* refers to processes and structures that allow people in the community to share perspectives or claim justice. Representation has both an internal and an external element: symbolic framing—being seen and addressed as a subject in a shared world, and democratic voice—being capable of making yourself heard and contributing in your own voice (Rosa, 2017).

We can connect this understanding of representation back to a decolonizing perspective and its impact on research practices. Symbolic framing is about having your perspectives and identity validated and valued in society. It is a way to counteract dominator culture and the coloniality

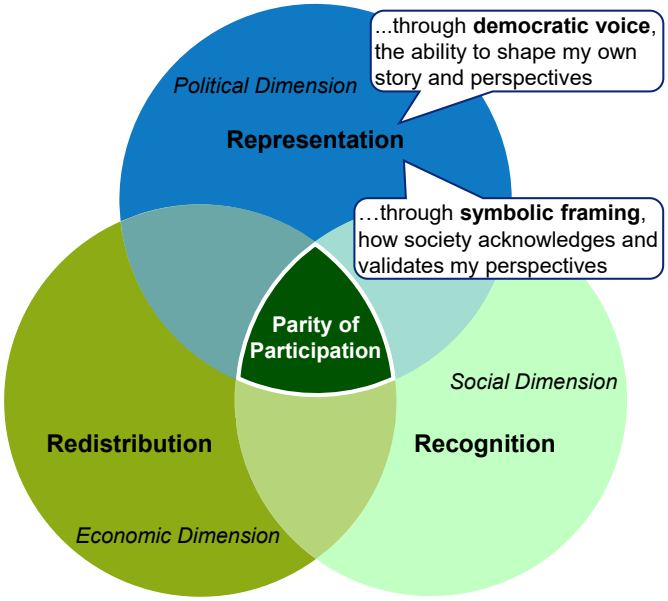


Figure 2.1: The prerequisites for parity of participation. In Fraser's framework for social justice, *redistribution* addresses inequality in the economic dimension (resources and opportunities), *recognition* addresses inequality in the social dimension (privilege and social status), and *representation* addresses inequality in the political dimension (contribution to discourse). This figure was adapted from Blue et al. (2019).

of power, which attempt to enforce a superior way of knowing through the subordination (or erasure) of other perspectives. Symbolic framing also refers to the power that researchers have to define a subject and influence our societal understanding of that subject. When researchers do not consider certain subjects, especially in data science, the lack of representation creates problems in methods, theories, and our capacity to address real-world challenges (Franklin et al., 2023).

While symbolic framing is about how others represent us, democratic voice is connected to how we understand and represent ourselves. Streck (2021) describes a concept from the Brazilian philosopher Paulo Freire which elaborates on the idea of democratic voice: "authentic word" is the possibility to express through words or actions your "human ontological vocation to be more"—i.e., the meaning or purpose for which we exist (2021, p. 91). Both democratic voice and authentic word describe a form of agency. They refer to our ability to act in ways aligned with our values and our ability to express our perspectives and understanding of the world.

Although agency is connected to our internal framing of self, it is also influenced by external conditions. Chin observed that researchers select topics based on their experiences, interests and perspectives, while also privileging questions that would gain 'scholarly attention' and fit in scope (Chin et al., 2022). There is a conflict in academia between our desire to pursue what we value and a need to conduct research that fits within the expectations (and acceptability) of academia. Representation in academia also looks like expanding what is considered acceptable to reduce this tension. In this way, participation (through representation) can also enable individuals to experience greater self-efficacy (Rosa, 2017).

Agency is only a small piece of Fraser's framework connected to how we share our stories and thereby represent our perspectives. However, I place additional focus on this concept because storytelling and acknowledging our positionality as researchers are decolonial acts that can unsettle the objectivity and decontextualization of subjective human experiences (Anthony-Stevens & Matsaw Jr, 2020). Smith writes that "the need to tell our stories remains the powerful imperative of a powerful form of resistance" (1999, p. 35). My aim is to evaluate a version of inclusion that facilitates efforts to decolonize academic spaces; I am interested then in looking at agency both as an indicator of representation and as a motivator for change in the community.

Meaningful inclusion, starting from parity of participation, is a combination of internal and external factors. Externally, inclusion depends on having access to the opportunities and resources

that allow for participation (**redistribution**) and being seen by others as possessing valuable and valid perspectives (**representation**). Internally, inclusion depends on individuals feeling capable and motivated to share their perspectives (**agency**) and able to acknowledge how social status is intertwined with and impacts other aspects of inclusion (**recognition**) (Figure 2.2).

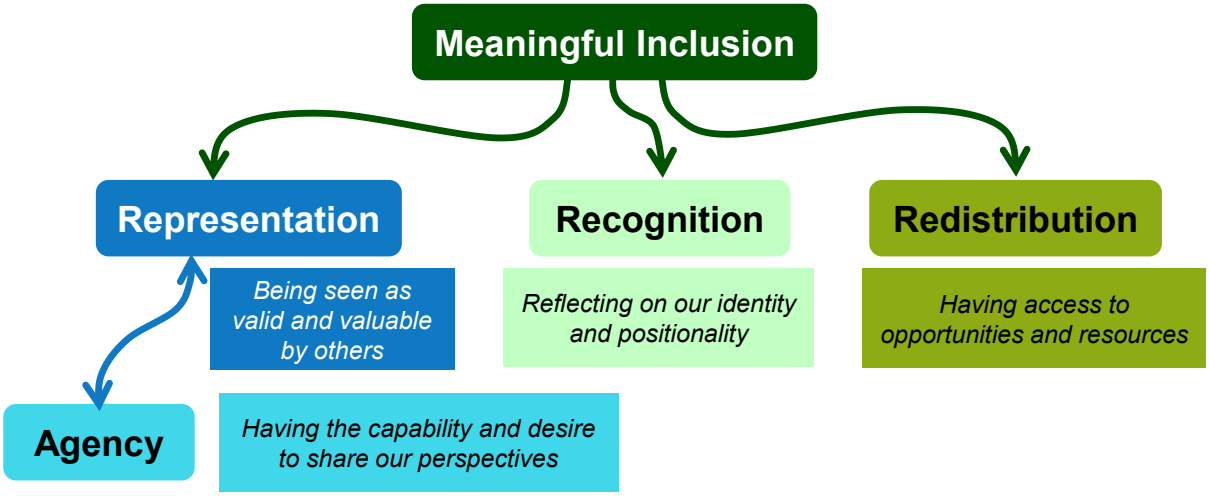


Figure 2.2: Conceptual framework for understanding meaningful inclusion. This is a simplified framework for understanding meaningful inclusion based on concepts from Nancy Fraser's Parity of Participation (Blue et al., 2019), but interpreted through a lens of decolonizing research practices. Specifically, I expand on Figure 2.1 in two ways: 1) I place emphasis on the role of agency in representation by including it as a separate concept. 2) I interpret recognition to be about the *internal process of reflecting* the privileges granted to us by our social statuses, rather than about the social statuses we possess. I interpreted recognition in this way to distinguish it from representation and because positionality and reflexivity are also prominent concepts in decolonizing literature (see Section 2.1.1 where I expand on these concepts).

2.1.1 Reflexivity and de-centering the dominant perspective

Falcón (2016) defines three feminist research principles which draw on a combination of transnational feminist scholarship and their own experience navigating geopolitical tensions between the U.S. and Peru as a researcher. These principles—*positionality and reflexivity*, *relational ontologies*, and *multilingualism*—add a layer of understanding to the internal and external factors that influence inclusion as parity of participation.

Positionality and reflexivity are about acknowledging our privilege in various contexts and challenging our internalized dominance (Falcón, 2016). The act of reflecting on our position is a form of recognition, because it allows us to consider the ways social status and identity shape our research and relationships. As Chin et al. (2022) explain, researcher positionality is underpinned by our identities, and this influences what we look for in research, how we understand it, and also how those we are researching see us.

Position is influenced by our identity, but it is also influenced by our relationships. An individual's positionality shifts and evolves as they engage with and relate to individuals, communities and institutions (Chin et al., 2022). This relational view of positionality is supported by Fish et al. (2022) who explain that social interactions shape our understanding of the world as we constantly try to construct meaning in various social contexts. Positionality both influences

and is influenced by our relationships, in part because those relationships can also shape our perspectives. For this reason, I consider positionality as a function of relationships to both people and processes in my research questions.

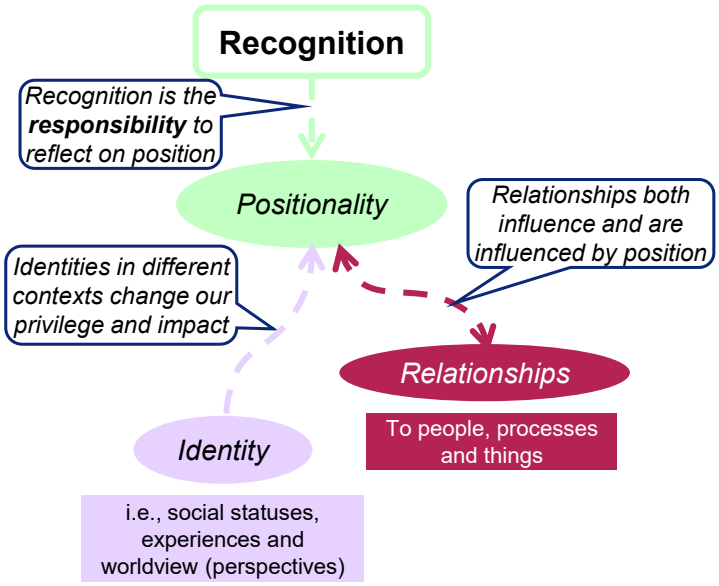


Figure 2.3: Expanded interpretation of recognition. This figure summarizes the connections between recognition and the principle of *positionality and reflexivity* described by Falcón (2016). The boxes below *identity* and *relationships* describe these concepts, while the speech bubbles describe the connections between concepts shown with dashed arrows.

Falcon's first principle also suggests that recognition in the community can be evaluated through positionality and reflexivity (Figure 2.3). In one of my co-creation sessions, a student explained that whether they feel included depends on how they see themselves in a crowd. This student also connected inclusion to a process of reflecting on their position relative to others.

The second principle of *relational ontologies* adds meaning to the concept of representation. Falcón (2016) explains this principle is about challenging the dualism embedded in Western research practices by acknowledging (and involving) multiple forms of knowledge. Bringing relational ontologies into our research also means accepting that not all ways of knowing can be researched empirically, and thus we as researchers must be creative about our approaches. Through this principle, Falcón challenges us to consider our willingness to step into uncertainty and believe that which we do not know and cannot explain. In my own process, this principle encouraged me to trust my intuition and rely on the experiences and insights of others in moments when I felt stuck. For the data science community, this concept of relational ontologies invites us to consider representation as how open the community is to the perspectives (and expertise) of others. It is also connected to agency, as we have to first feel like we can express our perspectives for them to be heard by the community (Figure 2.4).

Multilingualism also expands on the meaning of representation. When Falcón (2016) writes of the importance of multilingualism, they are challenging the dominance of English in our research practices. Language is a tool through which recognition and representation can be achieved or undermined. Musara et al. (2021) explain that Indigenous languages in South Africa are devalued in education systems in part because English is still considered the language of power, leading to a lack of recognition for these Indigenous groups and their languages. The students I discussed my framework with also emphasized that language affects inclusion at an international university like TU Delft. For one student, their way of speaking led to others having a biased perception of them. For the other student, language impacted their ability to connect with others and express their ideas. As Meital and Jason write, "those who do not speak English fluently or at all struggle to claim the right to speak" (Meital & Jason, 2022, p. 19). Language relates back to our agency through affecting our capability to express perspectives.

Disrupting the dominance of English through multilingualism is about more than just utilizing other languages. Falcón (2016) emphasizes that researchers are translators; we are tasked with navigating between ontology and language, with reconciling differences in how we know the world and how that reality is expressed. Applied to a community, the principle of multilingualism invites us to consider whether shared understandings in the community (ontology) are aligned with the expressed realities of the people in that community (language). This offers another way to evaluate representation by asking how well the community is reflecting the perspectives of individuals into communal practices.

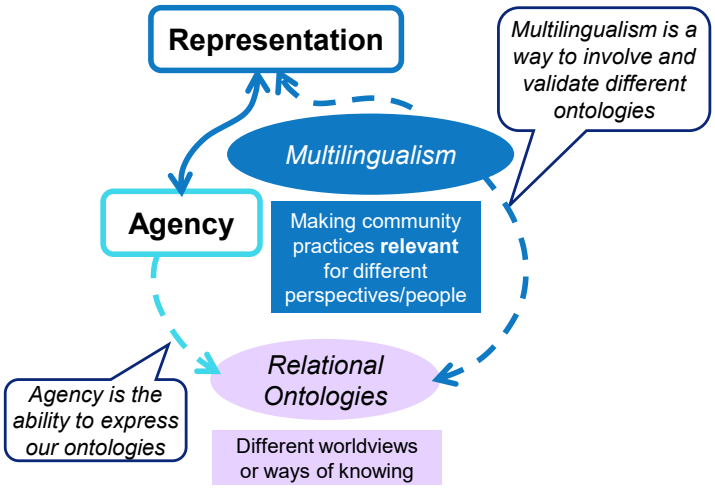


Figure 2.4: Expanded interpretation of representation. This figure summarizes the connections between representation and the principles of *relational ontologies* and *multilingualism* described by Falcón (2016), and *relevance* described by Kirkness and Barnhardt (2001). The boxes below concepts provide a definition, while the speech bubbles describe the connections between concepts visualized with dashed arrows. Multilingualism is interpreted in this framework as an action that can lead to greater representation and validate relational ontologies.

Inclusive spaces are spaces which challenge patterns of power and disempowerment, advantage and disadvantage (Musara et al., 2021). These transnational feminist principles are all ways to de-center the dominant perspective. We can create more space for underrepresented perspectives by recognizing and reflecting on our position, being open to the knowledge of others, and supporting individuals to express their reality in a way that is understood by the community.

2.1.2 Responsibility and inclusive relationships

Kirkness and Barnhardt (2001) use four words—*relevance*, *respect*, *reciprocity*, and *responsibility*—to call for a shift in the culture of higher education to one that is more inclusive for First Nations students. Reciprocity, respect and relationships are also key principles of Pacific research frameworks defined by Anae (2016) (as qtd in Fish et al., 2022). These words lend insight into how to characterize inclusive relationships, particularly between an institution (or a community) and individuals within that institution.

Relevance refers to whether universities construct their actions in a way that adapts to the needs of students, rather than enforcing a culture to which the students must adapt. Kirkness and Barnhardt (2001) write that institutional invulnerability is a mark of unresponsiveness; universities must be vulnerable to reflect and change. This is connected to Falcon's principle of multilingualism by expressing that institutions can (and should) act as translators between language and ontology. A measure of representation is a community's willingness to adapt, to change based on the needs of people in that community.

Kirkness and Barnhardt (2001) also call for universities to *respect* the cultural integrity of First Nations peoples and different ways of knowing. This form of respect is about acknowledging and legitimizing the knowledge, values, culture, and experiences that others have, much like

the principle of relational ontologies. On an interpersonal level, respect is also about having humility and an openness to learn from other perspectives. Evans et al. (2014) emphasize that building mutual respect and trust in relationships (especially with minorities or vulnerable populations) requires cultural humility and cultural safety. Humility is a way to provide the (external) representation that allows for cultural participation, because to have humility is to let go of the superiority of one way of knowing. As Jama (2023) writes, designing technology with a sense of humility gives space for the abundance of the world. Approaching relationships with humility gives space for the complexity of people.

Respect is also connected to *reciprocity* — the act of both giving and receiving in relationships. Kimmerer writes "Doing science with awe and humility is a powerful act of reciprocity with the more-than-human world" (2013, p. 252). Through our humility, we are giving respect—and in turn, we receive insights or learnings from those we interact with. Nathan et al. (2017) explain that learning from others is a way to engage respectfully. Similarly, Hampton (1988) judges respect between an individual and a group by looking at whether the relationship is seen as mutually empowering (as qtd in Kirkness and Barnhardt, 2001). Reciprocity in universities is about acknowledging that teaching and learning is a two-way process where students and professors can build upon the cultural contexts to which each are being introduced (Kirkness & Barnhardt, 2001).

Reciprocity is a quality of relationships that support inclusion. Fish et al. (2022) observed in their own attempts to involve Māori and Pacific communities in Aotearoa New Zealand that mutual interaction avoids non-participation which can lead to feelings of powerlessness or marginalization. They explained that reciprocity ensures that practices in the community are continually negotiated in relationship with others, so that people are not excluded from a space because their own cultural practices are in conflict with how the community operates. Falcón (2016) also connects reciprocity to the economic dimension of participatory parity, explaining that reciprocity demands concrete acts of redistribution. Reciprocity is about giving and receiving access to the resources that make it possible to be included. This is aligned with the experiences shared by the students who I talked with. They expressed that access to resources is limited because information about those resources is withheld or because access must be granted by people in gatekeeping roles (such as study counselors).

Finally, Kirkness and Barnhardt connect back to the idea of inclusion by describing responsibility through participation. Universities that truly wish to be more inclusive towards First Nations students must actively participate in creating respectful, reciprocal and relevant relationships with students (Kirkness & Barnhardt, 2001). Active participation is also a way for students to take responsibility for their own lives. Responsibility to participate and create an inclusive environment depends on both the individuals in the community and the conditions within the community. The internal and external aspects of parity in participation can also be aligned with individual and communal responsibilities respectively.

2.2 Applying a Framework for Meaningful Inclusion

Fraser's parity of participation provides a framework through which to define inclusion as participation. The concepts described by Falcón (2016) and Kirkness and Barnhardt (2001) suggest ideas to evaluate inclusion, particularly for underrepresented and non-dominant perspectives. Even just using the concepts presented in these three works, there are numerous angles through which I could assess inclusion. Given the scope of my project, I operationalize only a few parts

to answer my main research question (Section 1.2.1).

My first two sub-questions are about defining position in the community and determining how that position might impact participation. They attempt to evaluate external conditions—conditions within the community—that support (or prevent) inclusion. I look at positionality through our relationships because of the emphasis that both Falcón and Kirkness and Barnhardt place on the importance of relationships. I consider not only relationships between people, but also relationships to institutions (such as the university) and to processes. Each of these types of relationships falls within a different dimension of the parity of participation: to people (social), to material things (economic), and to larger systems (political or spiritual) (Rosa, 2017). To answer these questions, I ask people in the community to describe their relationships to others and consider their involvement, familiarity or access to various resources and processes in the community.

Representation is the other external condition I attempt to evaluate, though in a less straightforward manner. I ask whether people feel as though their values are shared by the community. The concepts of relevance and multilingualism suggest that representation is related to how a community can incorporate individual perspectives into communal practices or culture. Values can be both shaped by our individual worldviews and take on shared meaning in a community or organization. Although I don't specifically ask about changes in the community, in my analysis I also look for relevance through how people describe the community changing.

My third sub-question is also about representation and how identity shapes the way we express ourselves in community. To look at this internal aspect of inclusion, I ask about participant identity, values, and agency in research. I also ask about opportunities to reflect, which is a way to ask about recognition and to inquire if there is space in the community for reflexivity.

Figure 2.5 shows how I make concrete different aspects of the overall framework described in this chapter. The dark blue speech bubbles in this figure contain the actual concepts that I asked participants about in my data collection. Appendix C contains the specific questions that I prepared for interviews. There are a few concepts I did not ask about specifically, but that I still looked for in my analysis. For instance, I look for reciprocity and respect in how people describe their relationships with others. In the next chapter (Section 3.2) I will also elaborate in detail the process through which I conducted interviews with people in the community as well as mention my data analysis process. I also revisit the concepts from this framework to answer my fourth sub-question and design an intervention for the community (discussed in Chapter 5).

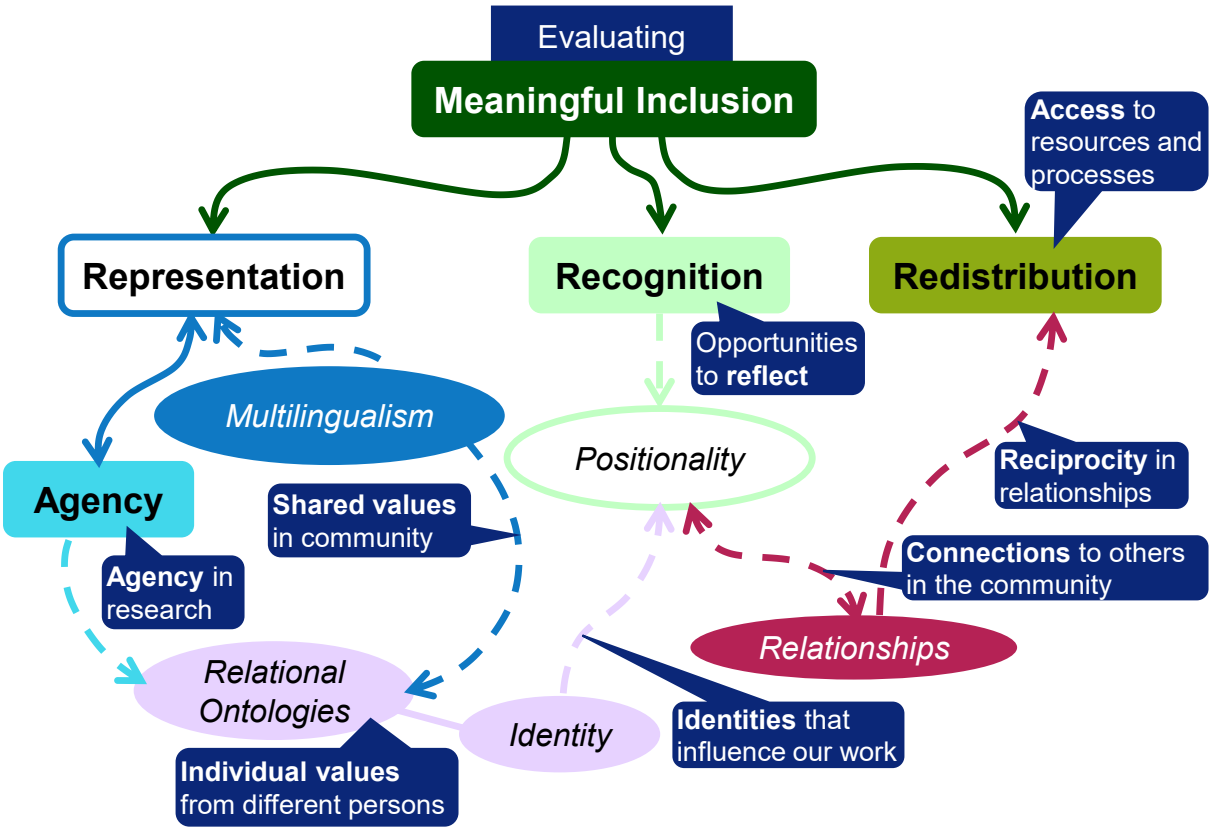


Figure 2.5: Evaluating meaningful inclusion through a combined conceptual framework. This figure provides a more complex view of meaningful inclusion that combines the interpretations of **recognition** (Figure 2.3) and **representation** (Figure 2.4) with parity of participation (Figure 2.2). The connections I made between concepts from different theories/principles/practices are shown with dashed arrows, while solid arrows indicate connections from the theory of participatory parity. Relational ontologies and identity are connected with a line because our ontologies (worldview) are a form of identity. In this visual, I also show how I operationalize these concepts and their connections. The dark blue speech bubbles contain the key ideas that I ask about in interviews and point to the concepts (or their related connections to other concepts) they are evaluating. Two concepts (representation and positionality) are shown as outlines, since I do not evaluate those parts of the framework directly.

3 | Methodology

In this chapter, I lay out the steps through which I conducted my research and design. Figure 3.1 shows an overview of the project: the first diamond was about learning from the community through exploratory research, which supported me in determining a design focus that I took into the second diamond. Along with describing what I did, I also reflect on the decisions and assumptions I made along the way. I highlight the areas where I attempted to incorporate decolonizing practices, and comment on instances when I also fell short. In writing this section, I am drawn to a recommendation proposed by Caxaj:

> "It may be important for critical and/or decolonizing researchers to map out how their analysis is action-informed and vice versa (how their action is analysis-informed) as a way to demonstrate how activities and events that are viewed as "unconventional" and outside of the research project, are in fact, central to the research process" (Caxaj, 2015, p. 10).

My thesis process was filled with serendipitous moments that shaped my approach in profound ways. I also share these seemingly "unconventional" details as a way to invite others to consider the influence of small actions and interactions in their own research processes.

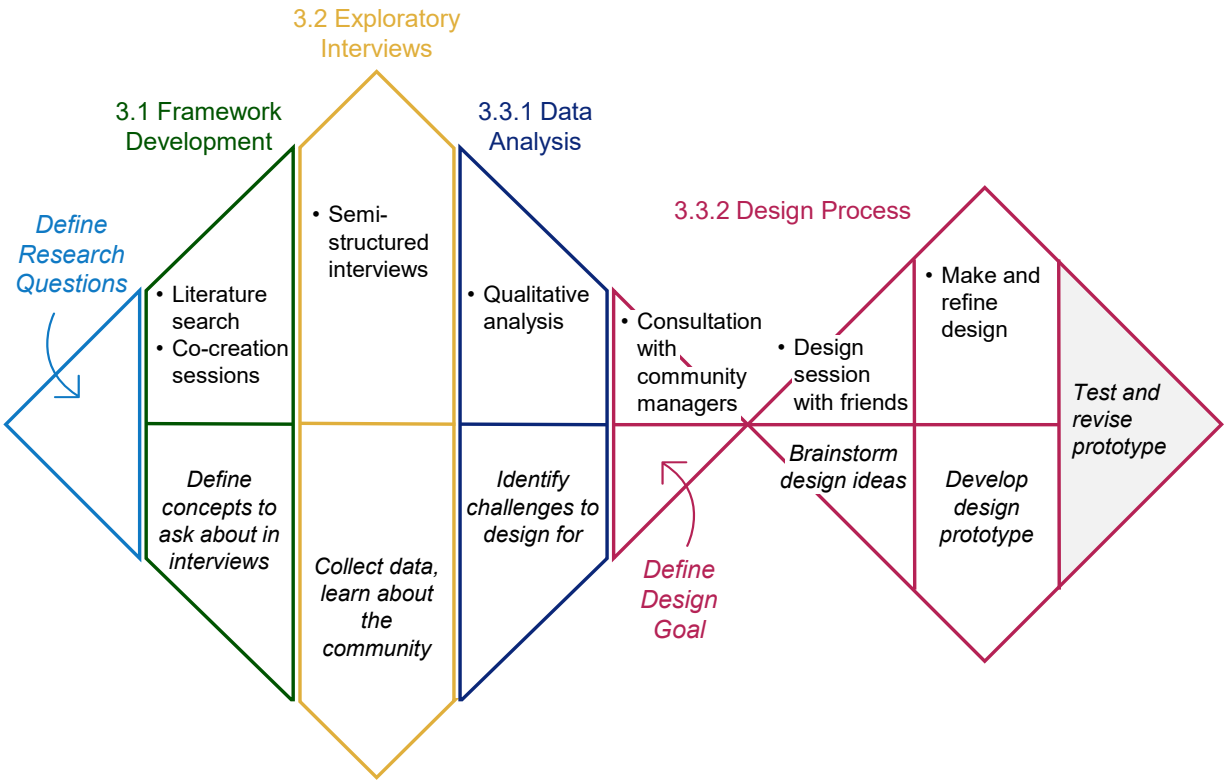


Figure 3.1: Overview of my research and design process. After defining my research questions, I had four stages in my process (labeled with the corresponding sections in this chapter). In the top half of the diagram, I list the different activities or processes I did in each step. The bottom half of the diagram contains the main goal or outcomes from each step. Although I did not have time to test and revise my design prototype (the last step in grey), it would be my next step in the overall process.

3.1 Framework Development

3.1.1 Constructing a Foundation from Literature

Starting from my research questions, I constructed a conceptual framework through which to understand meaningful inclusion from a decolonizing perspective. This project is about inclusion, so I made it a practice to include underrepresented voices from the very beginning of my process. Our values, beliefs and assumptions shape which literature we select and privilege in our research from the outset (Chin et al., 2022). As much as possible, I attempted to privilege those perspectives excluded from academic spaces. I intentionally sought out literature and scholarship from Black, Indigenous, Global South or other non-Western perspectives. I looked for literature that referenced or connected to books, poetry, and artwork created from these perspectives. Similarly, I found many of my sources by looking at the references of other articles, especially the articles that resonated with my own understanding of the world. I focused my framework on the concepts that appeared frequently and pursued literature from authors or scholars whose names recurred in many works on decolonizing methodologies.

I took a thematic approach to incorporating literature, as opposed to a systematic review. My goal was not to evaluate all decolonial methods, but rather to apply a perspective informed by decolonizing methodologies that might be relevant for the data science community at TU Delft. Below is a table summarizing the criteria I used to find literature, determine relevance, and prioritize which literature to read or include.

Table 3.1: Summary of literature search and inclusion criteria

Criterion	Included	Justification
Frameworks or Theories	• Decolonizing research, decolonizing methodologies • Indigenous knowledge, indigenous research methodologies • Feminist theories • Social justice theories	Besides a decolonizing perspective, I also include indigenous research and knowledge, feminist, and social justice theories. All of these perspectives are ultimately concerned with making visible and questioning unequal power distributions in society, so I found it relevant to draw on these frameworks, especially when referenced alongside decolonization.
Source format	All formats included	I intentionally chose not to exclude any source formats. The academic peer-review system is a form of gatekeeping and reinforces ways of knowing already privileged in academic spaces. Recognizing that experience is a valid form of knowledge, I also draw from art, blogs, poetry, books, and other narrative forms to build my framework.

Authorship	Preferentially included literature (co-)authored by Indigenous peoples, women, people of color, people from non-Western countries or institutions	Again, I am attempting to prioritize and give validity to perspectives historically excluded from academic spaces by citing them. I did not actively check the identity of authors, nor did I actively exclude literature based on the author. However, when considering which literature to read first or to skim for additional sources, I chose literature written by authors positioning themselves as from these identities.
Context	Literature about research or higher education (university and above). Did not include literature focused on primary or secondary education, medical or health systems, business, or politics	TU Delft is an institution of research and of higher education, and I selected literature that discussed a similar context. There are a few exceptions of articles that discuss reimagining education systems with examples from primary/secondary schools (see Fish et al., 2022; Musara et al., 2021). I still included them because they focus on participation in and implications for educational systems at large.
Databases	• General search engine • Google scholar • Web of Science • SCOPUS • TU Delft Library • Community resources lists	I started with typical databases to find articles. However, I also conducted general searches online to find sources that were not necessarily scholarly. I also drew on resource lists and community-created databases that suggested resources, including Indigenous AI Resources and the Alaska Native Knowledge Network.
Relevance	Sources discussing: • Decolonizing research or education practices • Data science and colonization • Feminist or social justice theories • Frameworks that resonate with or challenge my own understanding of the world	An aspect of relevance that I consider is the value literature brings to my own understanding of the world. Whether that is challenging me to reconsider my perspectives or reinforcing values I hold, I think it's important to consider the impact sources have on my own process and learning. This is a way I choose to use my voice in this thesis and represent my truth.

I experienced a kind of validation while conducting my literature review that is unlike any feeling I've had throughout my academic career. Despite possessing various privileged identities, there are also ways that I have been excluded from academia as well– whether that is as a woman in science, as a person of color at my predominantly white undergraduate institution, or now as an international in a country and culture that is sometimes at conflict with my own values. The literature I read for this project gives voice to the complexity that I possess. It emphasizes a humanizing form of research that feels aligned with my own values and perspectives towards relationships and people. I feel invited to speak my own truth and to reflect on my story and

how it shapes my academic discourse, without being questioned for why I am including these experiences. This feeling—when you read the words of others and think, *yes, that is how I understand the world as well*—this is a feeling I want to recreate with my project.

3.1.2 Co-creating a Framework

My worldview undeniably shaped the way I interpreted literature for my framework. Recognizing that I am coming from the US and working in the context of the Netherlands—both Western, privileged, colonizing powers—I felt that it was important to create a framework rooted in the experiences of people who are not coming from a place of historical power. I wanted to avoid positioning myself as an expert or authority figure with an inclination to teach and theorize (Eseonu & Duggan, 2022). It was important to me that I take the time to decenter myself and co-create with the communities I am attempting to serve with this research.

The main purpose of these co-creation sessions was to understand perceptions of inclusion and what decolonization might mean within the context of the TU Delft. While some of the literature I drew on for my framework was contextualized within higher education, I did not find examples of decolonizing efforts within universities in Europe. I suspect one reason decolonial perspectives have not been applied widely to European institutions is because colonization happened outside of this continent's borders. However, that does not mean that impacts of colonization do not reach the Netherlands—especially considering institutions like our own serve students from places still impacted by the coloniality of power.

To find people to participate in these sessions, I reached out to student associations at TU Delft that bring together people belonging to a particular region or culture of the world that is connected to Dutch colonial history. For associations in which I knew somebody, I reached out to my contact directly or asked someone I knew to pass along my request to their connection within the club. For organizations where I did not know anyone, I emailed the contact information provided on the association website.

Of the ten associations I contacted, two people were willing to discuss with me. These were both people I got in contact with directly, either because I knew them personally or because a friend introduced us. Although three associations responded to my email with an offer to pass my message along to their members, these did not result in any further communications. Appendix A contains a detailed explanation of the associations I included or excluded, the questions I asked in the co-creation sessions, and a summary of my notes from the conversations.

I had two fruitful discussions with international students that informed my framework and research design. There were a few important ideas that came up in these conversations that I carried forward into my framework and interview set-up. First was expanding upon a definition for decolonization: it must be an ongoing process and decolonizing efforts need to be concrete and actionable. Another concept that came up was related to resources: accessibility is limited because there is no public information available or because there are people in gatekeeping roles that control access. Finally, both students shared stories about experiencing racism in Delft, highlighting the importance of also talking about racial diversity. As one student put it, conversations about diversity at TU Delft are almost always about "white diversity" – by focusing on gender or nationalities within Europe, conversations about diversity neglect the role skin color plays in how people are perceived.

Input from these conversations led to three major decisions in my process. First was the decision

to focus on participation through resources and processes as a measure of inclusion. Second was the decision to counteract the dominant conversation about white diversity and instead scope based on whether participants identify as a person of color. And lastly, I set an intention to design something that would be actionable and have an ongoing contribution to the community.

3.2 Exploratory Interviews

The second step of my research was to learn about the data science community. This was the largest and most time-consuming step in my research. Especially since I would not consider myself a data scientist, I came into this project with very little knowledge about who is in this community or how it operates. I wanted to get a thorough understanding of how different people experience the community so that I could stay true to my design intentions. Placing emphasis on the exploration and data analysis phase of this project was necessary to define an actionable design goal that reflected the experiences, desires, and needs of the community.

I conducted semi-structured interviews to qualitatively explore my research questions. A huge portion of TU Delft is involved with data science, so to talk with a broad cross section of this community, I scoped for three criteria. First, I looked at two faculties to consider the influence of institutional culture. I reached out to people affiliated with the Faculty of Technology and Policy Management (TPM) and the Electrical Engineering, Mathematics and Computer Science Faculty (EEMCS). Both of these faculties have programs dedicated to educating students in data science but take different educational approaches. Second, I consider institutional position. I reached out to bachelor's and master's students, PhD candidates (PhDs), postdoctoral researchers (postdocs), and assistant and associate professors from both faculties. Third, as I am considering how identity impacts inclusion, I also attempted to talk with individuals from each category who do or do not identify as a person of color within the context of the Netherlands.

3.2.1 Recruiting Participants

Learning from my experience recruiting for the co-creation sessions, whenever possible I reached out to individuals through existing connections. I initially invited 31 professors through email to participate. This included seven assistant and five associate professors from EEMCS, and four assistant and twelve associate professors from TPM. Some of these contacts came directly from my supervisor who was familiar with researchers working in data science (primarily within TPM). To reach contacts in EEMCS, I first emailed a professor I knew personally from a non-academic organization we are both involved with. About a month into my data collection, I reached out a second time to professors who did not respond and invited an additional 7 associate professors from each faculty, since that was a group from which I had the least responses. Despite many professors declining to participate directly, a few helped indirectly by suggesting postdocs or other professors I could contact.

I reached PhDs in a similar way, asking two professors if anyone in their research group might be interested in talking with me. One limitation of this method to reach PhDs is that I exclusively talked with PhDs from two lab groups. This could have biased responses from PhDs as a category, as these participants share experiences from being in the same lab group.

To reach master's and bachelor's students, I visited classrooms in person and relied on my personal network. I gave a 2-minute explanation of my project and a QR code linking to a form through which students could leave their contact information and answer a few scoping questions

(Appendix B). I visited one course on data science methods for students in the Engineering and Policy Analysis Master's program (TPM), and two elective courses for students in the Computer Science Master's program (EEMCS). I additionally visited a project class for third year bachelor's students in TPM. I got four responses through this form, two of whom were people I knew previously from other courses. Two of the classroom visits also served as valuable opportunities for me to convince the professors teaching them to participate.

Finally, I relied on two friends in the Computer Science bachelor's program (EEMCS) to forward a message with the same form to a group chat for students in their program. These friends approached me in situations when I mentioned my project off-hand, curious and willing to help – I did not realize either were involved in computer science. Both are also friends for whom I have helped in the past, and I was reminded again of the importance of reciprocal and supportive community.

Towards the end of my interview process, I started getting more connections through others. I was put in contact with two postdocs that way. A few serendipitous moments also provided participants–a friend of a friend walking up while I was eating lunch with classmates, or a professor I exchanged emails with sitting in the office next to where I interviewed a PhD.

3.2.2 Interview Process

As with developing my framework, I attempted to approach my own data collection from a decolonial perspective. I framed these conversations as an interactive discussion rather than an interview, attempting to disrupt the dominant research model which distances academics from non-academics and encourages the privatization of knowledge (Falcón, 2016). These discussions were an opportunity to "actively learn through built relationality" (Chin et al., 2022, p. 29). I tried to listen and learn throughout the discussion, sharing my own experiences when relevant to find common ground and inviting participants to also share openly.

I also kept the format semi-structured. Although this made it more challenging to compare responses, it also allowed for more unexpected answers and the flexibility to adjust to the participant's needs. For instance, using a semi-structured format allowed me to adapt the discussion to be anywhere from 25 minutes to over an hour, or to be held in person or online. This flexibility also allowed me to construct meaning in the community throughout the entire interview process. Caxaj (2015) describes the practice of iteratively reporting back findings to participants to check understanding and increase overall meaning-making in the community. Although to a lesser extent, I was able to take some responses and perspectives from one interview into the next, providing more opportunities to put individual responses in conversation with each other.

The interviews themselves were split into three sections. First, I asked about how participants would define the data science community as a way to clarify the scope of my own work and guide the question. I then asked about what motivated them to do data science research, what they value as a researcher, and how identity impacts their work. These questions operationalize the internal aspects of parity of participation (identity, agency, and recognition) described in Chapter 2 (see Figure 2.5). In this first part, I also asked about whether participants felt their values were shared by the community, as a way to evaluate whether their perspectives were represented by others (the concept of multilingualism). The second part of the interview involved a drawing activity where participants were asked to draw how they are connected to the data science community. I asked questions about their drawings or invited them to explain what

they drew. This was a way to evaluate the role of relationships in our framework for meaningful inclusion by elucidating connections to others and following up with questions about how they might characterize different relationships. Finally, I asked about resources available to and processes happening in the community. I started with an initial list that expanded after each interview as participants suggested things that I missed. The list was also an open prompt for them to comment on what they felt was important, what they may or may not know about, and what they find irrelevant. This section addressed the concept of redistribution in the framework. The full interview protocol is described in Appendix C.

An ongoing commitment to knowledge sharing

An essential part of decolonizing research methods is the ongoing commitment to share knowledge from our research (Smith, 1999). This includes allowing for collective ownership of the data, analysis and presentation (Datta, 2018). I made sure to give participants a choice about how their data would be kept or stored for the future. After each discussion, I sent the completed transcript back to the participant so that they might review, revise, or keep it for however they might want to use it. This step was a way to add validity to my method by allowing people to confirm the perspectives that they shared with me. One participant was explicitly interested in keeping their own responses for reuse, and four participants revised parts of their transcript (primarily to improve anonymity).

The second way I committed to sharing knowledge occurred both as part of my interview protocol and as an informal process. The protocol produced a list of resources and processes that combined ideas from multiple conversations (Figure C.2). This provided an opportunity for participants to learn from others in the community, because they could see what was added to the list before them. Informally, I also learned a lot about both the participants I talked to and about processes at TU Delft. When I had the opportunity, I tried to connect the two. For instance, I shared a resource that I learned about in an interview with a friend. Conversely, I also shared an interesting article with a participant that was related to their research. This effort to share knowledge has extended into my personal life as well; especially when I learn about a resource or event that I was previously unfamiliar with, I consciously think who else might benefit from this information.

Streck (2021) considers the climate of trust and the relationships formed between researchers and stakeholders as a criterion of validity in decolonizing and transdisciplinary research. This metric was a central consideration in my process, informing even my smallest actions. It was important I did the little things, such as sending a thank you email when someone went out of their way to respond or bringing cookies to interviews as a way to give back and appreciate the time people gave to participate.

3.3 Analysis and Design

The last stage of my research was to analyze results from the exploratory interviews and use my findings to inform a design for the community. As with other stages of my research, I attempted to make this process collaborative whenever possible.

3.3.1 Data Analysis

After transcribing all of my discussions, I qualitatively analyzed responses by looking for themes, concepts from my framework, and relevant narratives. This was done in two rounds using ATLAS.ti software. First, I looked for codes that I already knew—such as concepts taken from my theoretical framework or answers I expected from the interview questions. Concurrently, I coded in-vivo (looking for new codes as I went through each document) that were thematically connected to my research questions. I highlighted recurring themes between conversations or stories I felt were interesting. In a second round, I then selectively coded for some of the themes that I identified, grouping similar ideas and then going back through all the documents to check for related quotes and stories from other participants.

The codes themselves were not as important for my analysis as the quotes that accompanied them. With the exception of coding for values, my analysis was purely qualitative and I did not keep track of the number of occurrences of each code. Instead, I used the codes to organize similar ideas from participants and guide a thematic construction of my results section.

Both the transcription and data analysis processes were very isolating. I feel as though I missed an opportunity to approach data analysis collaboratively. I considered on a few occasions involving my participants – for instance, sending them parts of the recording that were unclear to help with transcription or asking them which part of the conversation they felt was most interesting when I returned transcripts to them. However, I felt that these approaches demanded more of my participants. I did not feel comfortable asking more of my participants knowing that I had little to offer in return. Considering both their busyness and my own lack of time, I found it difficult to be truly interactive in my analysis.

3.3.2 Design Process

Moving from the analysis phase to designing was perhaps the most challenging step in my process. Ideally, my results would form the basis for brainstorming design ideas. However, thinking back to the quote from Caxaj at the beginning of this chapter, this was an instance when actions oriented my analysis more than analysis oriented my actions. I relied on intuition to take the initial step from results to design, and employed two techniques that help me in moments when I feel stuck. First was relying on art instead of words, and second was to talk to people.

In the first diverging step of the design process, I took quotes that highlighted limitations or complexities in the data science community. I then painted abstract ideas of how to address those complexities (a scanned version of this image is in Appendix D, Figure D.1). These images I showed to supervisors and peers, narrowing towards a few concepts that others found exciting or interesting. In this first converging step, I also reached out to two community managers – from the Open Science Community and the AI, Data, & Digitalization Community – to gain additional feedback on what they would find valuable in a design. Together, this input helped me narrow my design to focus on a specific challenge: how to get micro-communities (e.g., lab groups, centers, or other thematic communities) to interact and collaborate.

This challenge was still too broad to work with, so in a second diverging step, I again reached out to two friends involved in data science research or studies to join a collaborative brainstorming session. In this session, we focused primarily on how to actually tackle the challenge of getting micro-communities to interact. We considered potential physical interventions, dig-

ital interventions, and institutional changes. The results from this session are in Appendix D (Figure D.2). While this brainstorming session again brought out more ideas, it also helped me specify what exactly I was attempting to change about the community. I will discuss my design in greater detail in Chapter 5.

Although I was able to involve a few friends and mentors in my design process, I did not have the time to properly validate or test my design with the groups I was directly designing for. Given more time, I believe this design could benefit from more iterations, not only to narrow the scope of the challenge, but also to refine the design itself.

3.4 Ethical Considerations and Data Management

My main two ethical considerations in this project were protecting the identity and anonymity of participants and representing their experiences and feelings accurately. Especially given the ways that research has been a dehumanizing process in the past, I wanted to be careful with how I constructed a story around my responses. One reason I opted to use gender neutral pronouns throughout this thesis was to avoid misrepresenting individuals who did not share their gender identity. This was also an additional way to protect people's identity.

My research involved people and required collection of sensitive and personal data. This included names, contact information, and in some conversations, information regarding a participant's racial or ethnic origins, political opinions, or religious or philosophical beliefs. Participants also shared information about their research and relationships to others, which greatly increased the chance of them being identified by others in the community. In my project planning, I went through the Human Research Ethics Committee (HREC) approval process. This involved writing a detailed checklist of potential ethical concerns and creating a risk mitigation plan with actions to address these concerns. For instance, I took extra caution, especially in preparing the transcripts, to ensure anonymity. The fact that a few participants also revised their transcripts to remove potentially identifying information further highlights the importance my participants also place on data protection. As part of the HREC process, I also created a data management plan to store and protect the sensitive and personal data collected – both in the short and long term. My data management plan was prepared in collaboration with the data steward in the Faculty of Applied Sciences at TU Delft, and my HREC proposal was approved (a copy of the informed consent forms and risk mitigation plan submitted with my HREC proposal are in Appendix E).

4 | Results

4.1 Participants

In Section 3.2.1 I describe how I recruited individuals to participate in my project. Figure 4.1 visualizes the people that I reached out to and who agreed to an interview. The individuals who were most willing to help (either by participating or by connecting me to others) were typically people with whom I had already established some kind of relationship. This is especially evident in recruiting students. I visited four classes, each with over 20 students; yet only three students from these classes agreed to an interview. One of these students I had a conversation with directly after my visit, and a second student I previously met in a different course. Effectively, I only interviewed one student whom I did not know.

The response rate was also quite low from professors. Many professors did not respond at all, and about a third declined due to lack of time. Two professors, one from each faculty, declined because they felt they were not data scientists (despite that others in the community recommended I talk to them). PhDs and postdocs were a much more responsive group and the most willing group to participate. Almost all of the PhDs I emailed got back to me, and in total five agreed to participate (a sixth offered but felt they might be out of scope). Both postdocs had my email forwarded to them by a professor (whereas for all other PhDs/professors I sent an email personally). This may have also had an influence on their willingness to participate (see Section 4.5 for more discussion about the influence of supervisors).

In total, I conducted 18 interviews ranging from 25 minutes to an hour and 15 minutes. I talked with one associate professor, four assistant professors or lecturers, two postdocs, five PhDs, five master's students, and one bachelor's student. Eight of these people were affiliated with the faculty Electrical Engineering, Mathematics and Computer Science (EEMCS), and eleven were affiliated with the faculty Technology and Policy Management (TPM). One participant was involved with both faculties, so they are counted as part of both. However, for anonymity reasons, they are only shown as part of EEMCS in Table 4.1. Some participants were also part of other faculties outside this scope; in those instances I only consider their affiliation with either TPM or EEMCS.

Table 4.1: Participant position and faculty

	TPM	**EEMCS**	**Total**
Bachelor's student	0	1	1
Master's student	3	2	5
PhD candidate	3	2	5
Postdoctoral researcher	1	1	2
Professor	3	2	5
Total	10	8	18

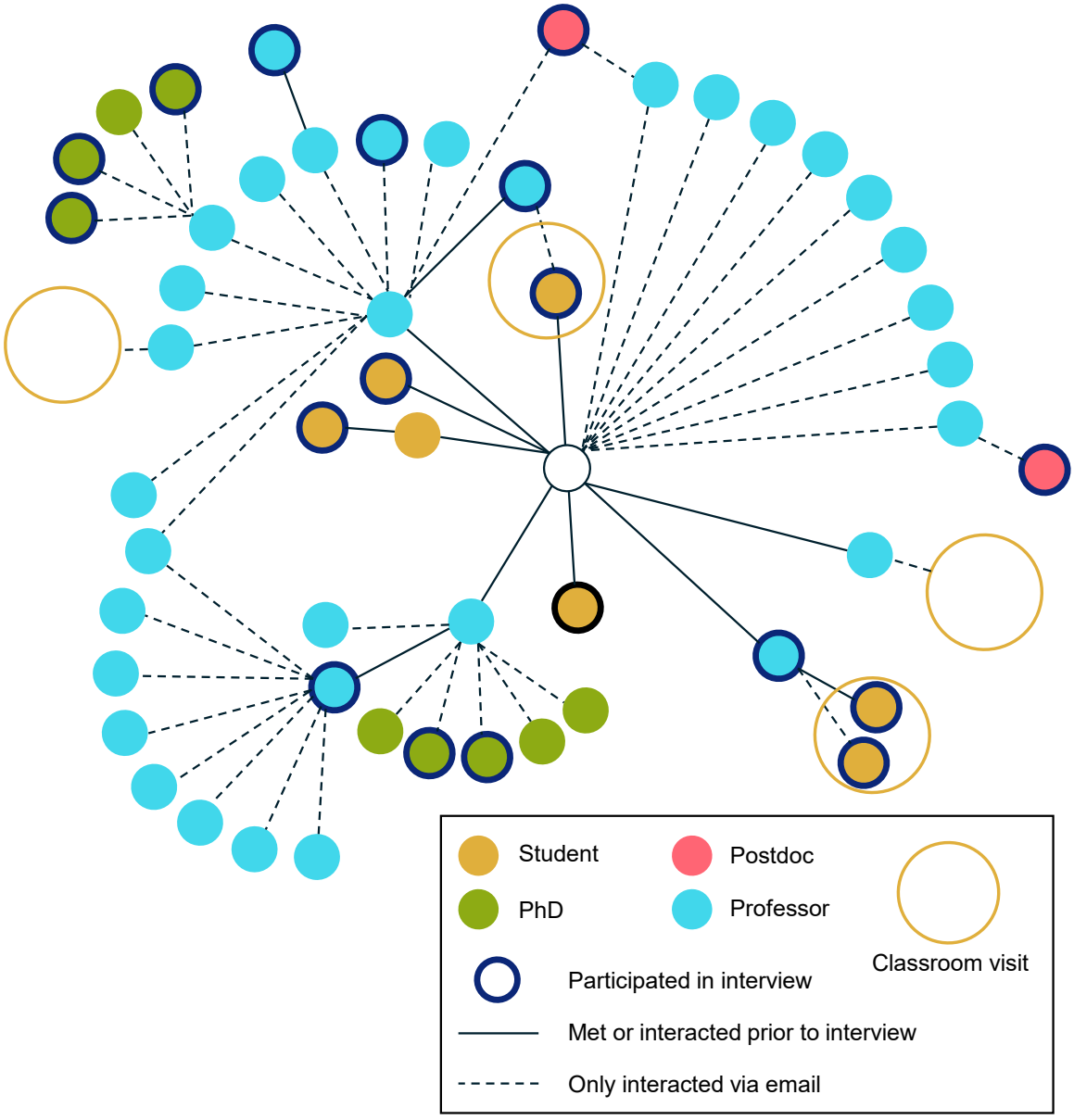


Figure 4.1: Connecting with (potential) participants. Each circle represents a person that I asked to participate; those who agreed to an interview have a dark outline around them. Some people helped by allowing me to present in their classroom (large yellow-outlined circles) or by suggesting names of others. These suggested people (or classes) are connected with lines to the people who recommended I contact them.

4.1.1 Scoping for Identity

I also considered whether someone identifies as a person of color (POC) or not as one of my criteria to scope participants. Some participants, especially students who were recruited through an interest form, provided this information upfront. For others, I asked if they identified as a person of color either during the interview or afterwards. In a few cases, this question was asked via email, and I did not always receive an answer.

Although I asked this question as a binary, in practice it produced a more nuanced answer than I anticipated. There were sometimes discrepancies between how people identified and how they were perceived (either by me or by others). One person acknowledged that while they are perceived as a person of color because they are visibly brown, they do not associate with

that label. Another person commented that although they recognize they are visibly white and from a Western country, they associate more with the identity of POC and would still consider themselves as such. Some participants were unclear with their answer, and instead interpreted the question to be about whether they were treated differently because of their appearance or POC status. One person said they didn't want to claim that label, because they felt it would be unfair to those who are treated differently. One person responded by claiming an identity that was neither white, nor POC.

Despite these nuances, using POC as a scoping criterion still allowed me to ensure I had a variety of perspectives in my results. Four participants did identify as a person of color, two did not associate with the label of POC but did not identify as white, two did not give a clear answer about how they identify, but instead explained that they did not feel as though they were treated differently because of their appearance, seven identified as white, and two were asked after the interview but provided no response.

4.1.2 Participants: Discussion

Although I was not surprised by the difficulty of getting professors to participate, I was surprised by the lack of willingness from students. This could be in part because especially students do not yet feel like they can call themselves data scientists; this was a feeling expressed by a few master's students I talked with. If these students do not feel like they are data scientists, I imagine they might also feel as though inclusion in data science is also not their responsibility. This was my first indication that perhaps students are not considered part of this community enough, an idea I will revisit in later sections.

I make some additions to participant scoping in Section 4.2.1. In my original scoping, I initially chose POC as a way to consider identity in part because of my co-creation sessions (Section 3.1.2). When we talk about diversity, it's important that we talk about racial and ethnic diversity. Appearance was an important factor that came up in those conversations, because it was connected to those students' experiences of discrimination. Most conversations about diversity and inclusion at TU Delft are focused on white diversity (i.e. people with backgrounds from different European countries, gender, or sexuality) which misses a lot of perspectives. It especially misses perspectives coming from parts of the world that are non-Western and not given the same resources or attention in academia (see Chapter 1). However, in practice, asking whether people identified as a POC within the context of the Netherlands created a complexity I am not surprised by, but did not anticipate. In part because the language of POC is common in the US, I expected others would already know how to respond to this question. But my assumptions were challenged by my participants who pointed out that this term is not used as often in other parts of the world.

That said, I also could relate to the difficulty of identifying as "either / or." I resonate with the challenges of having a fluid and often ambiguous racial identity. Given that responses to this question were so ambiguous, I do not make any conclusions based on POC as a criterion. I only reference race or ethnicity in my results if it came up in other aspects of the conversation, such as when asking about identity or relationships, etc. To further protect the anonymity of individuals, when I do describe results connected to a different racial or ethnic identity, I avoid simultaneously discussing that person's institutional position or faculty. Humans are complex, and I want to hold space for that complexity and ambiguity while writing my results.

4.2 Defining the Data Science Community

I started this project with the intention to understand inclusion in the Data Science Community at TU Delft. However, how to draw boundaries around this community was not such a simple task. With the help of my participants, I attempted to define two questions: first, who is a data scientist? And second, who is a part of the data science community?

These are questions that many of my participants were curious about as well. It came up before interviews when people read through my project brief. One professor described asking these same questions when considering who to invite to an educational event on data science. As this professor aptly described, "is it the science of data? Is it the data of science? Based on how you would frame this there are people that can be included or excluded." I found this observation to be true even in my own process of finding data scientists to interview. Two professors initially declined to participate, stating that they did not consider themselves data scientists.

There were two moments during interviews where I asked people to explicitly reflect on how they would define the data science community. Towards the beginning of each interview, I asked participants in general terms who they would consider to be part of this community. Towards the end of the interview, participants were prompted to draw on paper how they are connected to the data science community. Responses to these questions varied widely. Some were specific, naming individual people or individual lab groups. Some were broad and general, thinking more about the structure of the community rather than direct or interpersonal relationships. Despite the variation in describing the data science community at TU Delft, there were a few trends in how people thought about the organization. These I have combined into five community representations which are ways of thinking about the structure of the community (Figure 4.2). Two of these representations are about defining who is a data scientist, and the other three are related to the question of who to include in the data science community and how to draw boundaries around and within it.

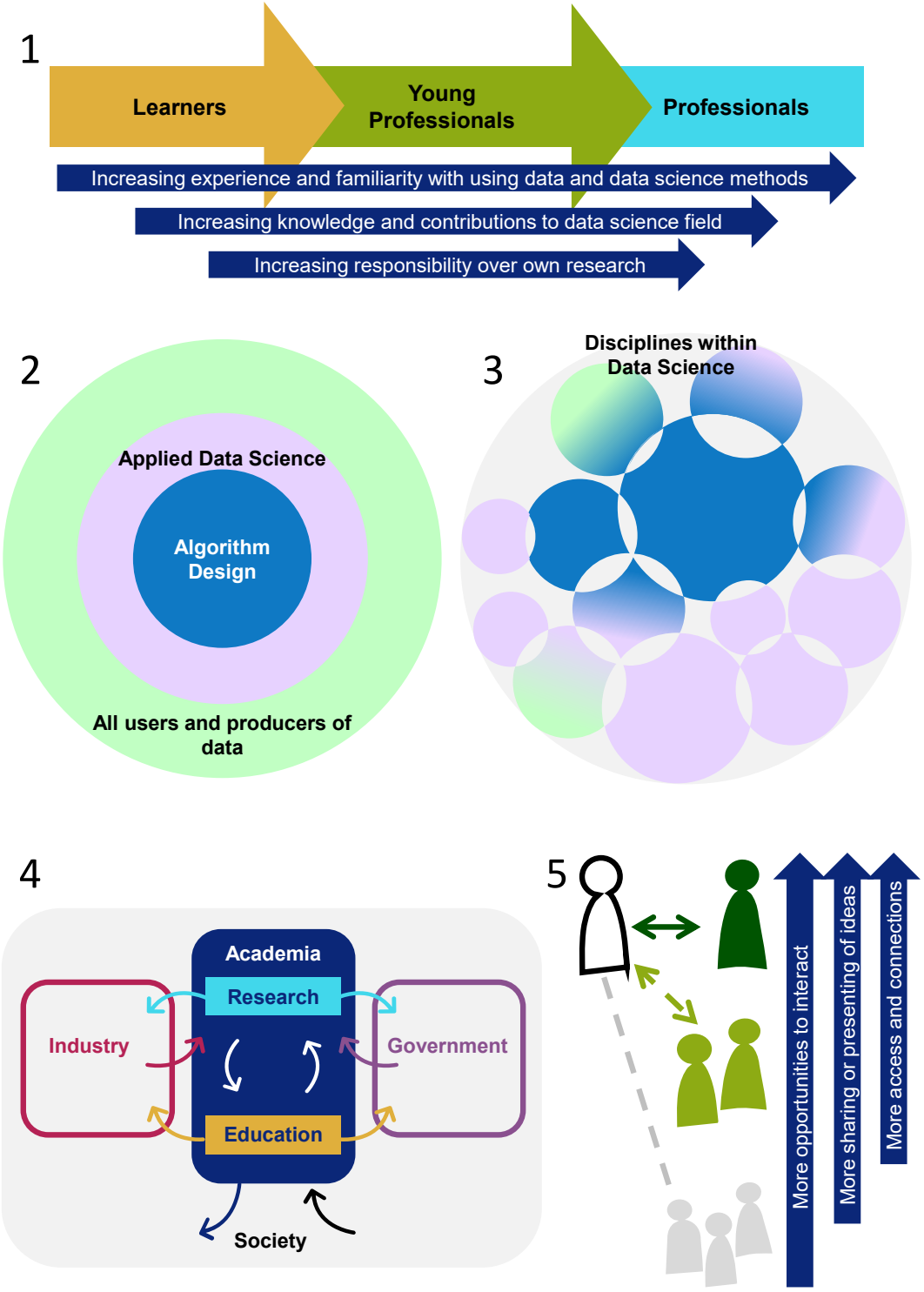


Figure 4.2: Representations of the data science community. Five ways of abstractly visualizing the structure of the data science community. Visuals 1 and 2 are ways to define who is a data scientist: 1) based on levels of expertise (Section 4.2.1); or 2) depending on how data is used in research (Section 4.2.1). Visuals 3-5 are ways of looking at interactions in the community: 3) the community is made up of different domains or disciplines that intersect (Section 4.2.2); 4) the community is made up of interacting institutions; or 5) the community is composed of interpersonal relationships (Section 4.2.2).

4.2.1 Defining a Data Scientist

"Everyone is in data, everyone is in algorithms, so I think most of us are to some extent involved in the data science field. But there is still a major called data science, data analysis. Now I don't know how do we define them."

This was a sentiment shared by many. At its broadest definition, a data scientist can be anyone who works with data. Throughout the interview process, I started to build my own definition, one which was challenged and expanded on by each subsequent interview. There were two defining aspects that emerged: the relationship between an individual and data, and their level of expertise in interacting with data.

Relationship to Data

The most common way that people defined who is a data scientist and thus part of the data science community was based on how people relate to or interact with data itself. One person in EEMCS described it as levels (Figure 4.3): there is the algorithmic level, which includes people working on fundamental algorithms or tools for data science in a generalizable context. Then there is the application level, which includes researchers who apply tools to real data. And finally, there might also be a third, broader level that encompasses all people who collect, use, or work with data (including myself).

These level distinctions were a definition I shared in later interviews, to challenge and see whether people agreed upon this distinction regarding who to include. In response, a few people also mentioned data management as a relevant but separate layer to consider in the community. Some also challenged whether to include all of these levels, or which level was most appropriately called "data science." However overall, there was quite a bit of consensus regarding how to split the levels themselves across all positions and both faculties.

Level 3: Everyone is a data scientist
Most of my interviewees acknowledged that "data science is a very broad concept." One student described data scientists as "everyone who does something with data, either for their work or studies." A PhD mused that maybe we could include "everyone in the world, because everyone. . . they are working with data actually." A professor also hinted that the amount of data might also be a relevant factor in determining who is a data scientist, because they would include "people that use data in copious amounts or generate data."

Since I am looking within academia, almost everyone I talked to was involved in research. Despite this, the broadest definition of data scientist still often extended beyond research uses. Ethicists and data stewards came up as potential members of the community. One student discussed that everyday people who use Fitbits or similar technology are also generating data. However, they felt that if you "can access data but not use it," that alone is not enough to make a data scientist.

Level 2: Applied data scientists
There are many ways to use data: collecting, analyzing, generating, processing, visualizing, and making predictive models – the difference distinguishing the broadest level from an applied data scientist is the intention behind use. This intention was described as "for the purpose of solving or understanding problems," ""to make conclusions out of data," or "for studying specific phenomena" within the data itself. As one student put it, "the data scientist is curious about the insights of the data itself."

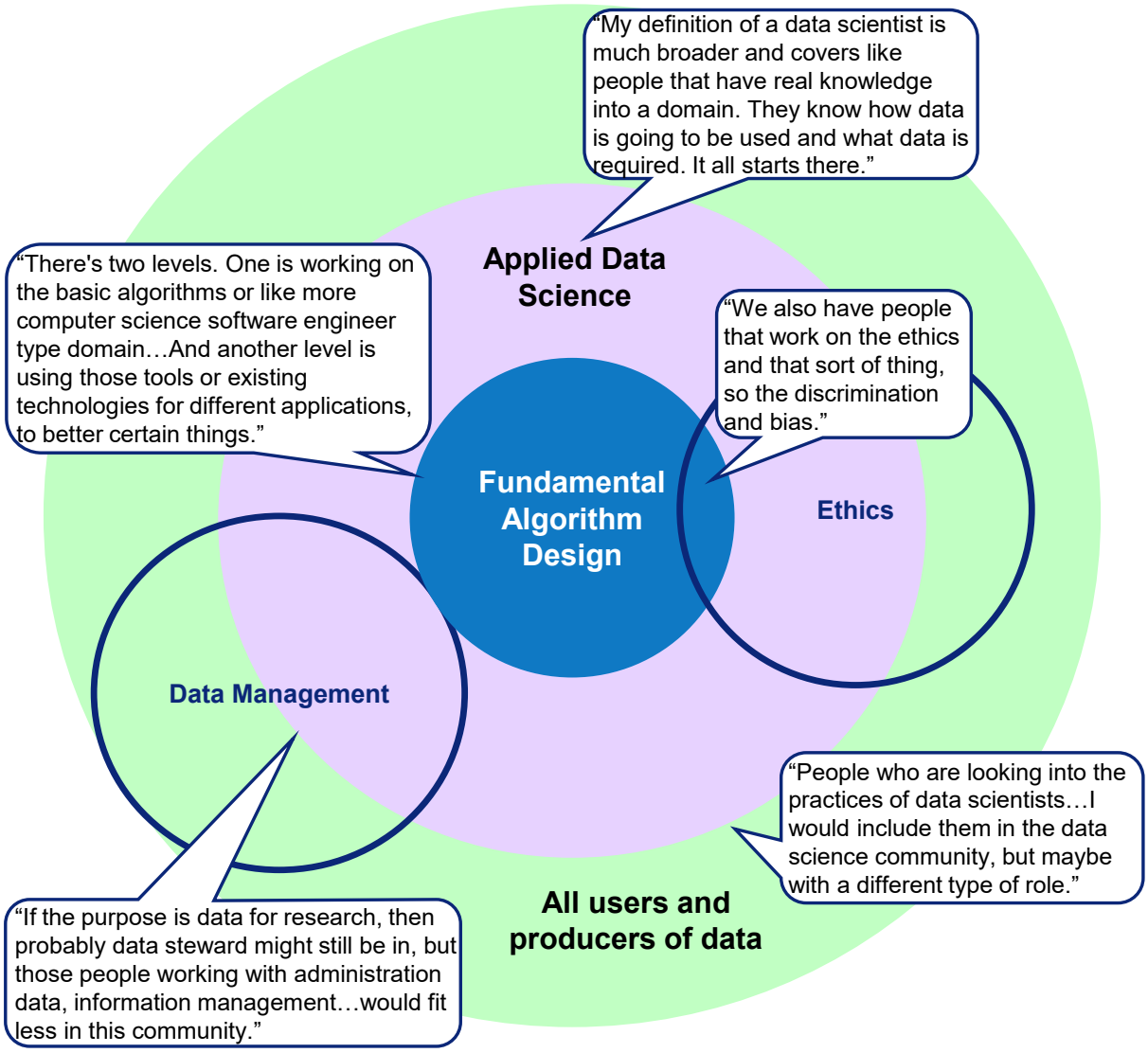


Figure 4.3: Levels of Data Science. A representation of who is a data scientist based on how they relate to or use data. Level 1 (innermost circle) is people who focus on moving forward algorithms or data science methods. Level 2 (purple circle) is people who apply data science methods to specific domains or look for conclusions within the data itself. Level 3 (outermost circle) encompasses all people who collect, process, produce, or otherwise use data. Most participants felt that the inner two circles encompass the data science community. It was less clear whether to include people who work on ethics in data science or data management, so these users of data intersect multiple levels.

Applied data scientists were often described as using tools and methods of data science. A professor explained this group contains people who "basically apply algorithms to datasets, and in their specific fields they're the experts." Another professor compared this group to the computer science community, explaining that applied data scientists "focus on making models represent the data as good as possible" – in essence, starting from real-world data to make models. One PhD made an additional distinction about the starting point of applied methods. There are data-driven methods that are guided by what is in the data itself, and there are model or theory-driven methods that are guided by background theories about the type of data that influence how you use it in your methods.

Level 1: Algorithm creators

While the applied aspects of data science were mentioned in almost every conversation, the

fundamental level of data science was mostly discussed by people affiliated with EEMCS or who had a computer science background. This level was described as "foundation work." It includes people interested in "moving more general algorithms forward." One professor described their work as "designing algorithms and theory. . . [that are] not attached to a particular domain of data."

Consensus-building: But who is a data scientist?
No one challenged the idea that applied data scientists belong to the data science community. However, it was less clear whether level 1 algorithm creators also belonged. Although there was consensus about who is in this level and their relationship to data – at least amongst professors – there was less consensus about whether algorithm creators should be considered "data scientists." The professor who initially did not consider themselves to be a data scientist instead placed themselves in the "computer science community" in EEMCS. For them, computer science and data science were two distinct groups. However, a professor in TPM with a background in computer science, highlighted that these two groups are often convoluted: "the connotation I hear more about data science, data scientist, is generally related to computer science." This was demonstrated by another TPM professor who had the impression that people from EEMCS "think of themselves as like the true data scientists." A third professor in TPM explained, "I understand that algorithms and new techniques are also part of this, but I think the majority of the people that are working on data science, they do more applied, you know?"

There is a perceived lack of consensus between professors in TPM doing applied data science and the computer scientists in EEMCS. However, the professors I talked to in EEMCS agreed that people with a real knowledge of their domain and understanding of the type of data they work with are the "true" data scientists. This disconnect was not evident in conversations with PhDs, many of whom felt that they were bridging the fundamental and applied levels of data science.

There was also disagreement about whether to include ethicists and people involved in data management. One professor discussed that they work with ethicists, but don't consider them data scientists. Another professor similarly would include in the community "people that work on the ethics and that sort of thing, so the discrimination and bias," however they did not specify if they consider this group data scientists. They did separate from both the applied and fundamental levels the "TPM faculty, that's more the broader implications of the this technology."

A few people also discussed data management as an aspect of data science – storing data, sharing data, collecting data, and cleaning data. One PhD explained they might include people performing these tasks in the community, because management has an impact on the data being used by people in the field. However, they would not necessarily call this data science. Another PhD explained that they believe all data scientists should be data stewards and manage their data appropriately, but perhaps not all data stewards should be considered data scientists. One professor had perhaps the clearest criteria of who to include from data management spheres, differentiating between data management that is administrative in purpose versus data management that is connected to research projects.

With the exception of whether to include ethicists and data management roles, there was generally consensus about the distinction between fundamental and applied – and from my perspective, both are equally important to the data science community. Even excluding level 3, this still encompasses an incredibly large group of researchers at TU Delft.

Level of Expertise

For a few people, the knowledge or expertise of an individual was the most important distinction in how they understood the boundaries of who is and is not part of the data science community. Both a master's student in TPM and a professor in EEMCS independently distinguished between three levels of expertise (Figure 4.4). The first level contained senior researchers or professionals, people that are highly respected because they have a lot of experience with working with data. Beyond just understanding data science methods, this group consists of people who also understand the social implications of their work. It was unclear what amount of experience warranted inclusion in this level, but the student described this group as "actual mythical creatures of like full-fledged professionals."

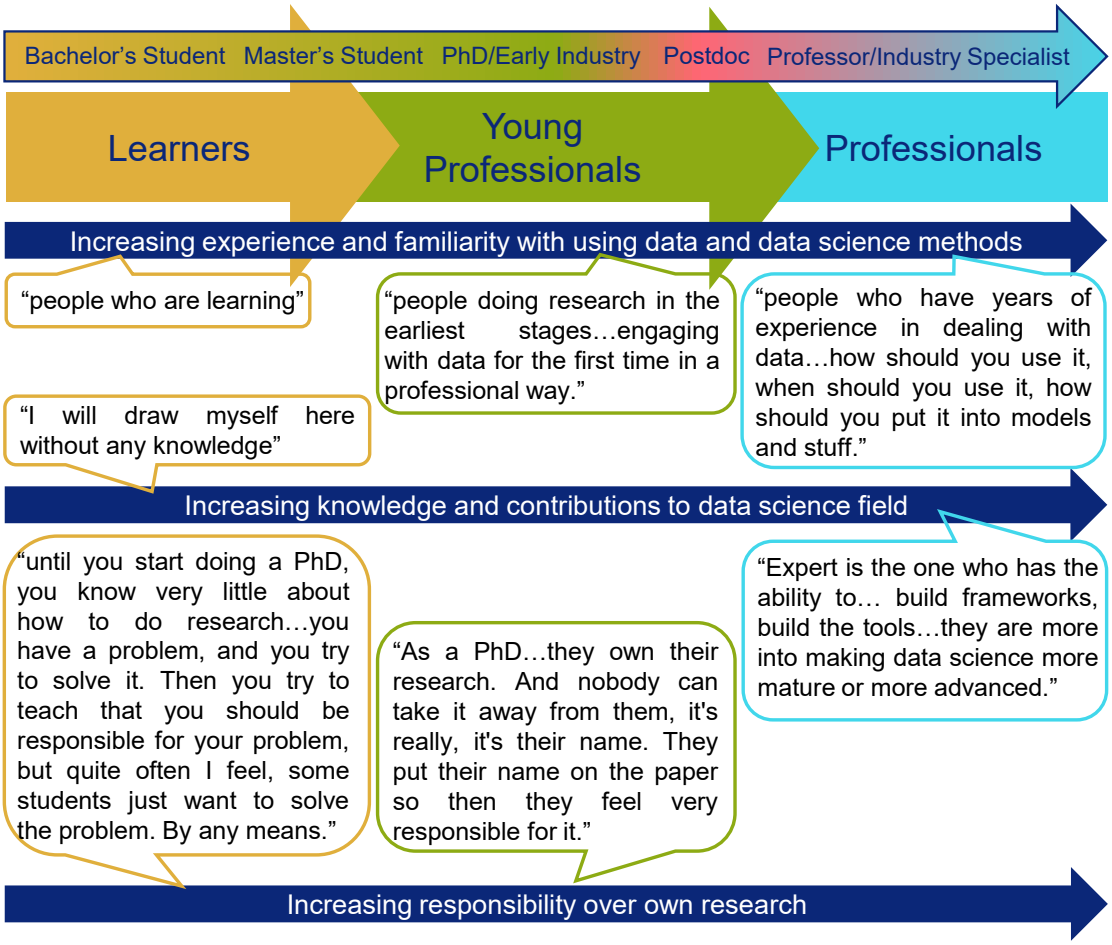


Figure 4.4: Levels of data science expertise. A representation of who is a data scientist based on their experience and familiarity working with data or using data science methods. The dark blue arrows indicate characteristics that participants felt made data scientists "more expert."

The second group was described as "young researchers" or "young professionals" – still data scientists, but not yet at an "expert" level. Both participants were unsure exactly who to include in this group, but they both connected this level to having greater agency in research. The student described this group as individuals "engaging with data for the first time in a professional way, so maybe master's students doing their thesis or PhD students starting their research." The professor explained their reason for this distinction is that PhD level students have greater ownership over their research, and therefore feel more responsibility for their work.

This was in contrast to students or the "learners" group. This group consisted of bachelor's and

master's students just starting their data science education, with relatively little exposure to data science or data science methods. They are also still learning the values and best practices of the field. Learners, while still included in the data science community, were not considered as "data scientists" yet by multiple participants.

While only two people explicitly structured the community around expertise, these levels were implicitly corroborated in other interviews when people described their own positioning. One master's student in their first year called themselves a "wannabe data scientist, wannabe." Another first year master's student considered themselves currently "without any knowledge" and separate from "the expert. . . who are deeply involved in the data science." Both of these students placed themselves in the "learner" group, not considering themselves as data scientists yet. One PhD also described themselves as "part of the young crowd," explaining that "in the academic ladder, PhD is the lowest, before studying of course. But then it's lower than postdoc, etc."

Extending Participant Scoping

In the rest of my analysis, I will also use these two systems for grouping data scientists to explore the experiences of people in different parts of the community. In some instances, I use these groupings in place of my original scoping (position, faculty) to protect the anonymity of individuals.

Table 4.2 shows these alternative groupings based on participants' relationship to data in their research. Six participants were not doing research in levels 1 or 2 when I talked with them and are not included in this table. Three researchers from EEMCS explained they don't work much with data beyond using it as a benchmark to validate their methodologies and are considered algorithm creators. Three PhDs who described their work as both developing generalizable methods and applying those methods to a specific domain are grouped as both level 1 and 2. Six researchers are considered applied data scientists and because they focus primarily on what they can learn from or do with data in their domain.

Table 4.2: Participant relationships to data in their research

	TPM	EEMCS	Total
Level 1: Algorithm Creators	0	3	3
Level 2: Applied Data Scientists	6	0	6
Both Level 1 and Level 2	2	1	3
Total	8	4	12

Table 4.3 shows groupings based on expertise. I included in the "learners" group the one bachelor's student and two master's students who have not yet started research for their theses. In young professionals, I combined the PhD candidates, two master's students who were more than halfway finished with their thesis research when I talked with them, and one master's student who worked in industry as a data scientist for multiple years and thus has experience engaging with data science in a professional way. Professionals included all professors and postdocs.

Table 4.3: Participant level of data science expertise by faculty

	TPM	EEMCS	Total
Learners	1	2	3

Young Professionals	5	3	8
Professionals	4	3	7
Total	10	8	18

4.2.2 Defining Community Boundaries

When defining a community, "what is important to me is to define the purpose of this community." This comment was shared by a professor who is involved in multiple smaller communities or initiatives connected to data science, each with its own motivating purpose. For this professor, a shared intention determined the boundaries of interaction. In this section, I explore how other participants visualized and described their interactions with the community.

Drawing Connections to Understand Community Purpose

Not everyone knew where to start when I asked them to draw how they are connected to the data science community. I encouraged them to consider drawing research groups, specific people, faculties, or whatever else felt relevant. Some started with my suggestions, but others chose to start with different types of connections. The differing connections participants choose to include in their drawings demonstrate a variety of purposes they have for engaging with this community.

Three people started by drawing their research and how their research process connects them to the data science community. This group included two of the three algorithm creators and one PhD whose research is both fundamental and applied. As one professor explained, "actually talking to you, I understand more the connection" between their research and data science. This professor did not initially consider themselves to be part of the data science community, and so for them, their research is what connects them to other data scientists.

For two students in EEMCS, they started by drawing how the community could be a resource for them to start a company. Reflecting on the activity, one student explained "as I was doing this, I was mentioning like if I was starting a startup, who would I ask first?" This student included interpersonal relationships that they felt could grant them access to the resources or support necessary to form a startup. The second student was more explicit about drawing their relationship to resources. They included representations of funding, human resources, data, and also the TU Delft as a source of knowledge. In both of these examples, the purpose of the TU Delft data science community is to support their own expansion of knowledge and resources for their current or future endeavors.

There were two other students from TPM that also focused on knowledge and their connection to knowledge. However, rather than thinking in terms of future start-ups and industry, they visualized how they understood the landscape of data science knowledge and their relationship to varying levels of expertise. For them, the purpose was also expansion of knowledge, but for the sake of learning and becoming more expert themselves.

Three young professionals did not include themselves at all in their drawings without prompting. Instead, they visualized relationships between institutional structures, positions, and data. One participant drew the overlapping roles of data steward and researcher, one drew three pillars of industry, government, and academia, and the third focused on the three levels of data science research within academia described in Section 4.2.1. These young professionals looked

at community from a macro-scale, considering first the institutions that data scientists are a part of before considering who those data scientists might be.

While these participants didn't think to include themselves at all, one student organized their drawing entirely around the familiarity of their personal connections. They explained, "I think this is the more relevant thing. Ok, how smart you are, it's good if you are smart and you can find problems. But if you don't have any connections, then what's the point?" There were other participants who also found it easiest to start by drawing specific relationships with other data scientists. They explained as they drew, "I think the easiest for me is first to place like the researchers from the EEMCS department" or "I will put the close collaborators, mainly outside TU Delft." A few other participants added in these relationships later, after thinking about institutions or lab groups. Some drawings listed specific names; others included more general roles. Nearly half of the drawings included direct relationships with people in the community. Relationships to others is another way participants considered community organization and purpose.

In addition to personal connections, many of the drawings also included overlapping circles to represent the lab groups, domains, interest groups, faculties, or institutions that participants are a part of at TU Delft. These can also be considered smaller communities within the larger umbrella of the data science community, each with their own distinct purpose.

I group these various purposes into three representations that give insight into the structure of the data science community: as a combination of institutions, as a network of relationships, and as a collection of micro-communities. These representations are also useful for contextualizing other aspects of my conversations with participants, as almost everyone switched between thinking about the community on a macro-scale to considering it on an interpersonal level.

Institutional Community

An institutional perspective of the data science community came up both in drawings and in conversations. Instead of thinking about people, this perspective considers the larger institutions or organizations that are part of the community. Companies, government, academia, and the public were all mentioned as having different relationships and importance to the overarching data science community. Figure 4.5 visually summarizes how participants described the interactions between these institutions

Companies were included in the community as collaborators, resources, or as a potential influencing force. Two students explained companies are important because they provide data. A professor similarly explained that companies are "releasing tools and data sets, and also they have their own research teams," which can influence the community as well. One PhD included companies because they knew many PhDs at TU Delft are doing their research in collaboration with a company.

Government did not come up as often as companies and was only included in drawings from people affiliated with TPM. Two young professionals discussed government as collaborators, because they both worked on research projects connected to government organizations or regulations. A few participants from both faculties also discussed government in connection to changing data regulations and data protection for both research and industry. A student clearly distinguished between industry companies and government by labeling them as private and public institutions respectively.

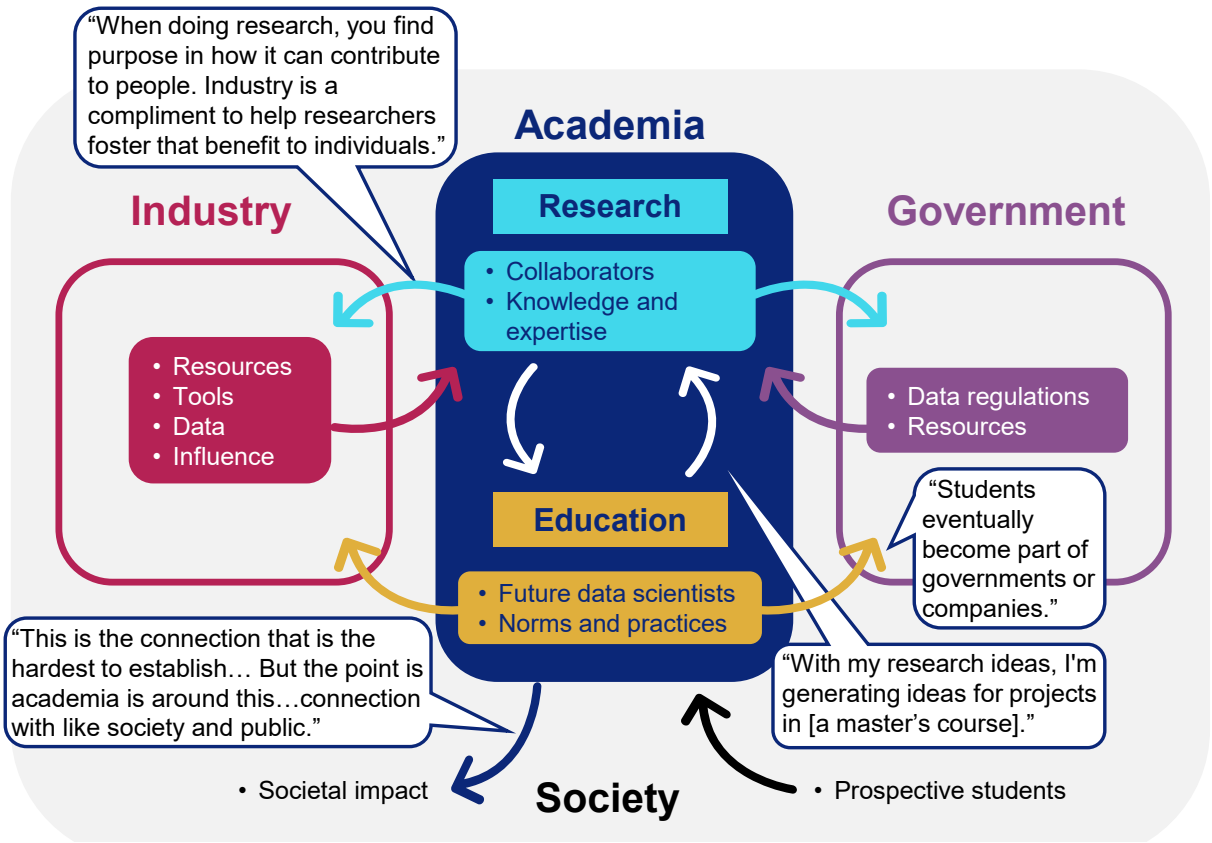


Figure 4.5: Interactions between institutions in the data science community. A way of representing the data science community based on institutions (academia, industry, government, and society) and how they interact. Arrows indicate the kinds of exchanges that participants described as happening between these institutions.

A few people also mentioned the "public" in connection to the data science community. One PhD specifically included "hobbyists" who have an interest in data science but are not formally involved in either industry or academia. For others though, the public was not necessarily involved in data science, but rather the group that data science should be serving with our research. This group was also referred to as "society," "home" or "friends and family."

Finally, academia was discussed as its own entity as well. One student described TU Delft as the "powerhouse of ideas." Another student mentioned professors may have relationships with "partner universities." Academia's role was more versatile than the other institutions described, which may be a bias of only talking to people within academia. It was included as a resource for knowledge, as an institution that provided education, as a location for research, and as an important connector between the other institutions mentioned. For example, a PhD connected society to academia through companies, explaining that interactions with "industry is a compliment to help researchers foster that benefit to individuals." I explore in more detail the position of academia in bridging institutions in Section 4.7.

Relational Community

A relational view of the community focused on the people in the community rather than structures. Participants discussed both the roles of other people and their relationship to those people. I also asked in a few conversations how they might characterize these different relationships (see Section 4.5 for more detail).

For one master's student, the familiarity of social connections was the main factor around which they visualized their community. They drew three layers: "close friends," "acquaintances," and "random people" separated based on how easy (and how likely) it is for this student to talk with them (Figure 4.6). Arrows and lines were another way that familiarity was visualized. One participant explained that the length of arrow indicated closeness, while a second participant explained that whether the line was solid or dashed represented how established that connection felt. The direction of arrows also indicated whether they felt a relationship worked in one direction or two.

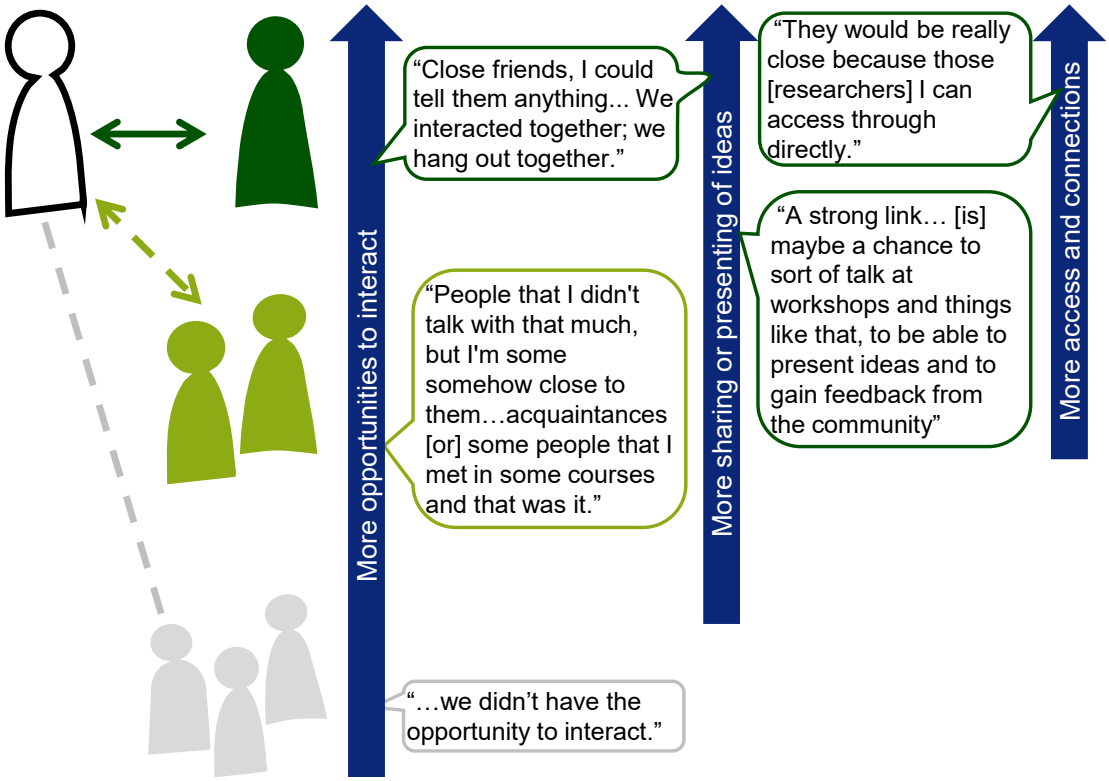


Figure 4.6: Closeness of relationships in the data science community. Another way to organize the data science community is by the closeness or strength of interpersonal relationships. The vertical blue arrows highlight characteristics mentioned by participants that indicate a "strong link" or "close" relationship.

Participants also indicated various roles people have in their drawings such as: students, teachers, professors, researchers, supervisors, fellow PhDs or classmates, and collaborators they work with. Participants who drew relationships with specific people or roles mostly included individuals either in the same position or in a position above them in level of expertise, but not those who were at a lower level. For instance, three students included other students in their drawings, and two of these students also included professors or teachers in their drawings. On the reverse side, only one professor drew students in connection with their teaching responsibilities. Both this professor and one student indicated that this relationship is one way and represented it with a unidirectional arrow. This same student however drew the relationship between students as a bidirectional arrow, indicating they can mutually "influence." Three professionals and one PhD only included peers, research collaborators or their supervisors in their drawings. None of the professors drew students or PhDs they were supervising. There were only two people who included positions both above and below them in level of expertise; a PhD and a postdoc who are both in mentorship roles included students they supervise as well as their own supervisors.

Disciplinary Community

Between an interpersonal and an institutional way of thinking is a representation of the community based on research domains, lab groups, or other micro-communities. As one PhD explained, "I also feel like something that's interesting about the data science community is people that come from kind of different fields can kind of easily get a little bit involved with data science." There were three professionals and one PhD from TPM who drew specific groups they interact with. They drew their lab groups, domains relevant to their research or interests (e.g., AI, transportation, climate change, etc.), and faculties within TU Delft.

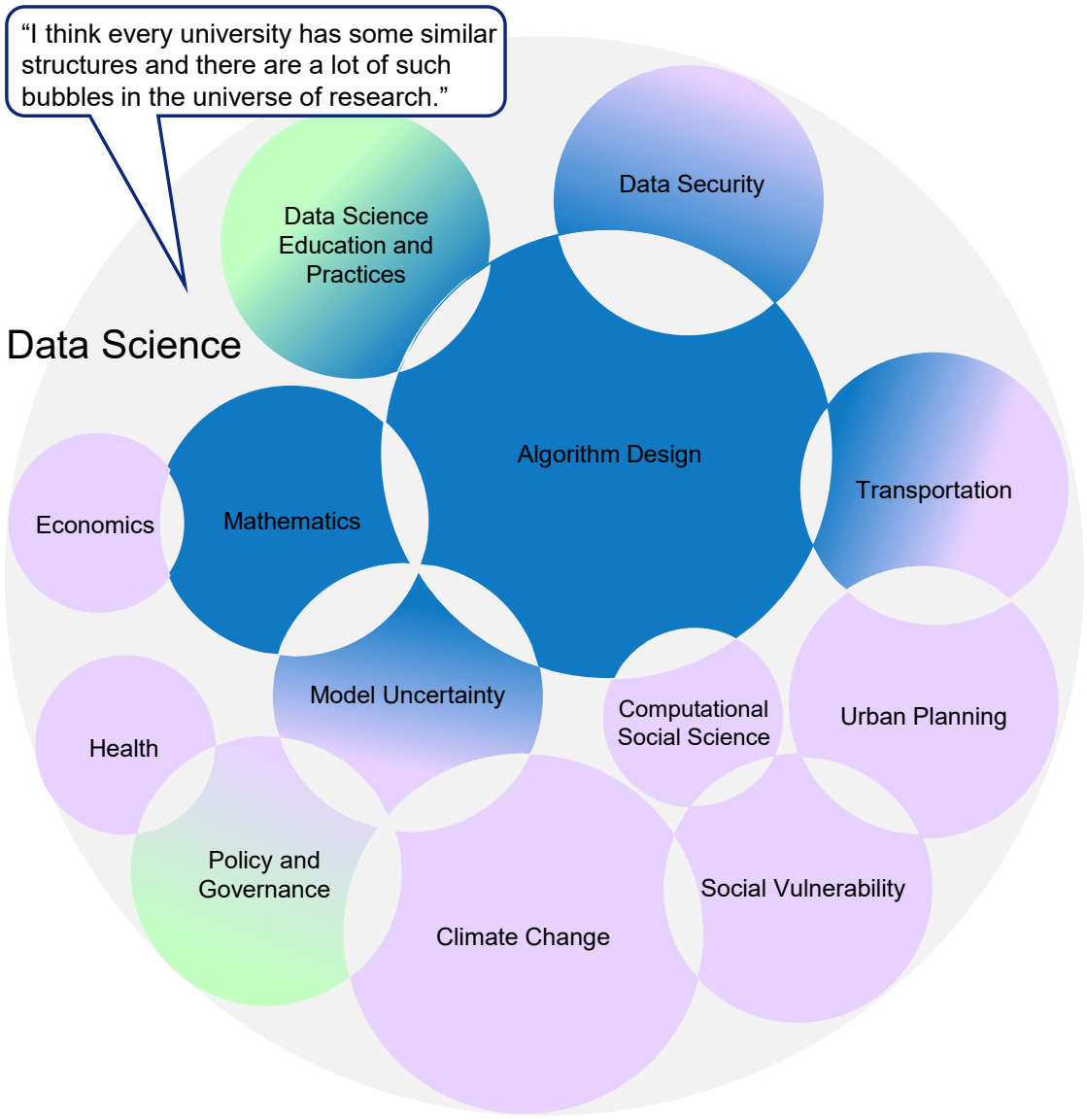


Figure 4.7: Diverse and interconnected disciplines within data science. This figure is an abstraction meant to demonstrate the diversity of disciplinary connections within data science research. I categorized the research that participants are working on based on how they described it to me in our conversations. These categories do not necessarily reflect how participants would describe their own research and vary in specificity. Circles intersect where one (or more) participant is doing research or has a background in both areas. Circles are sized relative to the number of people whose research I considered to fall within that category. The colors correspond to Figure 4.3 and whether the people who are doing research in that area are working on fundamental algorithm design, applied data science, or broader practices and uses of data (or a combination).

One justification for why it might be valuable to think of the data science community in this framing was given by a professor in EEMCS: "I feel you have to develop algorithms for specific domains and not apply deep learning everywhere out of the box." This professor also included various domains in their drawing, but in an abstract way rather than as specific groups or micro-communities.

From conversations, it was apparent that many of the researchers I talked with intersect or collaborate across multiple disciplines or fields. Figure 4.7 is an abstracted visualization of the intersecting and diverse domains within data science. Participants also drew distinctions between disciplines when they mentioned their educational background. However, some of the distinctions were not always so clear to me. For instance, one participant explained to me that university programs usually separate machine learning methods from traditional optimization based on mathematics, despite the fact that both are used in data science. A student in EEMCS similarly explained that TU Delft has two distinct tracks within the computer science master's for "Data Science & Technology" and "Artificial Intelligence Technology." However, AI again often gets convoluted with both machine learning and data science. As a professor in TPM explained, "in practice I would associate data science more with computer scientists and the people that do machine learning or AI." Another participant from TPM similarly grouped theoretical computer scientists with mathematicians and physicists, distinct from both "engineering research" and from data policy and regulation. They affiliated computer scientists with the EEMCS faculty and data policy with TPM.

This association between certain disciplines and specific faculties demonstrates how a disciplinary perspective of the community is tied to existing structures within the university. A disciplinary perspective is helpful in highlighting the challenges and opportunities data scientists face in working across or between disciplines, especially within the faculty structure of the university. Section 4.4 and Section 4.7 will elaborate on these challenges further.

4.2.3 Community Organization: Discussion

Some of the representations that I describe in this section might seem obvious. For instance, I think most people intuitively know that having more interactions with a person indicates a closer relationship. However, defining who "belongs" to a community has implications for how a community might be inclusive or not. We have to first acknowledge who we think should be in a community before we can assess whether those people are actually included or involved.

There are three disconnects that I want to highlight from this section. Rereading how various participants described applied data scientists versus fundamental data scientists (those working on or creating algorithms), there appears to be a disconnect between the two faculties in my scope. The professors I talked with from TPM had an impression that computer scientists in EEMCS thought of themselves as "true" data scientists. While in fact, some of these algorithm creators see themselves as adjacent to the data science community. On the other hand, at least one professor in EEMCS conflated TPM with ethicists and interested in broader societal implications. Their comment suggests that TPM is not nearly as well known for its more practical or applied data science research. An impact of having separate and siloed faculties is that it can lead to misunderstandings (or preconceptions) about what researchers are doing in each faculty. These preconceptions in turn might lead researchers to not consider people in one of these faculties as important for the community, despite that they are involved with data science.

Thinking in terms of disciplines can lead to similar misconceptions. A community manager for AI, Data & Digitalization was quick to point out to me that researchers under this umbrella don't like the term "discipline." Especially when it comes to areas of research within data science like AI, machine learning, etc., it is unclear where one ends and another begins. Datta (2018) describes disciplines as subcultures with their own language, concepts, tools and experts. Yet many of the disciplines that could fall under the umbrella of data science share concepts, tools and experts. At the same time, data from different domains also come with its own concepts and language. Smith (1999) explains that disciplinary approaches in research allow scholars to distance themselves from others and absolves them of responsibility for things that occur in other disciplines. Almost all of the data scientists I talked to intersect multiple disciplines. Perhaps this is also a reason for the strong sense of responsibility within the community; to be involved with data science is to inherently challenge the detachment accompanying disciplinary approaches (I highlight community interpretations of responsibility in Section 4.3.1).

I also noticed a disconnect occurring on an interpersonal level, specifically between students and others in the community. None of the master's or bachelor's students I talked with included specific people, groups, or domains in their drawings, despite that some are involved in lab groups and work directly with supervisors or researchers in data science. Meanwhile, PhDs and professionals drew all of these types of relationships. That no students thought to draw direct connections to people indicates that perhaps they feel less connected to the community, even when they are. At least when it comes to tangible relationships, students are very much on the periphery.

I will refer back to these various community representations in other sections of this chapter.

4.3 Values

Values, in their broadest sense, are an indication of what we believe is important. In asking about people's values, my goal was to gain insight into what the community believes is important and how that varies between positions, faculties and identities. In each interview, I explicitly asked participants about what values they have in their work or as a data scientist. I followed this question by asking whether they felt their values were reflected or shared by the community. In this section, I start by summarizing the values mentioned by community members and their interpretations. I then look at whether these values are shared both numerically and qualitatively based on how participants answered the second question.

4.3.1 Types of Values

One challenge of asking about values in an open-ended way is that people interpret values differently. I ended up using the questions participants asked me to also help clarify and explain what I meant by values in other interviews. For example, one participant interpreted values to be about valorization, and explained to me why their research methods and data had added value to society. One participant interpreted values to be about the characteristics or personality traits they possess. A few participants asked me what I meant by values, and in these instances, they provided their own interpretations. For example, a few asked if values included: "ethics", "professional or personal values", "moral or design values", or if it meant the "higher guideline of my research." These interpretations will also guide the summary of the types of values I present in this section. These include:

- **Personal values:** character traits or individual attitudes that people find important for either data science or for interacting with or collaborating with others in data science.
- **Method values:** these are instrumental values related to the process of doing research. These are things that researchers felt were important in how they approach their research.
- **Design values:** these are characteristics that people felt were important to incorporate into their research. They are less about the approach to research, and more about the qualities the data, methodologies, or outcomes of research possess.
- **Value as meaning:** a few people interpreted values to be about the meaning that their research provides. This includes values connected to the purpose, goals, or ideals people have for their research.

I identified values based on how they were explicitly named or if they implicitly tracked with similar definitions given by others. Some definitions I interpreted a bit more generally, combining examples from the interviews to encapsulate an overall meaning for certain values. I also look at how often certain values are mentioned and by which participants. Although I look at how many participants mention certain values, it is important to remember that just because someone did not mention a value does not mean that they don't find it important or agree with that value. Counting how often a value is mentioned is only a rough way to assess whether there are certain values that are more talked about or widely shared by these researchers.

Personal Values

Some participants considered values to mean the character traits or attitudes they felt were important for data scientists (or researchers more generally) to possess. For instance, two participants discussed the importance of **perseverance** and not giving up in research: "you have to pursue again and again and again, not to just withdraw from the first bad results." One person also referred to this determined mentality as being **ambitious**. A similar value was being **rigorous** to "make sure everything is taken into account" when creating your models. Although mentioned in connection with validity—a value more related to the research process—rigor was described more as an attitude of being thorough in the research process.

One participant also described perseverance as not feeling disheartened when you get feedback. A different participant described a similar idea as being **open-minded**. They described this as being "open to that feedback and also to the experience of other people, even if they don't resemble yours." An open-minded mindset was about receiving feedback, considering perspectives different from yours and accepting potentially unexpected opportunities in life.

Some personal values mentioned were more about interacting with others than individual attitudes. **Respect** was a value that participants both tried to embody and that they valued in their interactions. Specifically, they felt it was important to be respected for your time, the decisions you make and "for the work that you've done and are doing." **Empathy** was a related value that was implicitly discussed. One master's student mentioned that they respected data scientists who demonstrate empathy: "not only are you a professional, but you are a professional that cares about people." A few professors also expressed that paying attention to an individual's well-being is an important part of forming relationships that influence the learning process, supervision, and inclusion in data science.

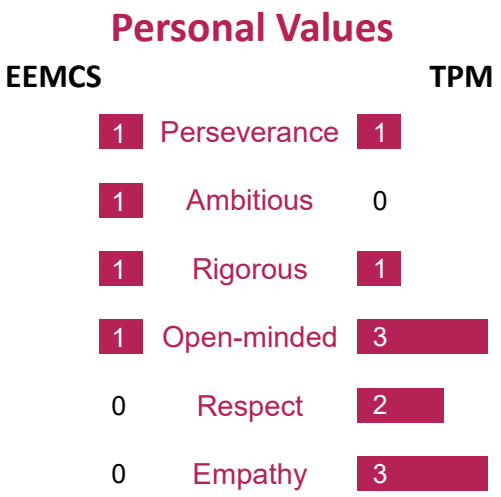


Figure 4.8: Personal values: number of mentions per faculty. The number of participants from each faculty who mentioned a particular value in conversation, either implicitly or explicitly. A participant is counted only once, regardless of how many times they discussed a particular value.

Figure 4.8 shows how many participants from each faculty mentioned these personal values. Overall, personal values were not talked about as frequently as method or design values. Although people from all levels of expertise mentioned these values, master's students seemed more likely to interpret values to be also about personal characteristics. Collectively they named more distinct values in this category than in any of the other categories.

Method Values

Many of the values people brought up in conversation were about considerations for the process of doing work as a researcher or data scientist. While some were more general, others were relevant to research specifically with data. Figure 4.9 shows an overview of how many participants mentioned different method values, divided by faculty.

Privacy and **data integrity** were both named as important values related to how researchers

treat data. Privacy was described as both protecting data and also having respect for people's differing attitudes towards data collection and use. Data integrity was explained by one participant as "making sure that you approach that data with respect in some ways. Because it is data that belongs to or has been extracted and is being used for different purposes." This value is also connected to a responsibility to be knowledgeable about where your data come from.

Transparency was another value that could be more generally applied to research but was also described as especially important for data scientists. A few people explained transparency as being "upfront about what are your modeling assumptions". However, others built on this definition and explained that being transparent as a researcher (or a decision-maker) meant being "explicit each step of the way and having a sound justification" for your decisions, asking "why did you make certain choices," and communicating both the message of your research and the limitations. Transparency was also connected to inequality by a participant who explained that "transparency is important... because then you can actually have a conversation about something serious and not just push everything you know to the side, start hiding things. Because when you hide stuff that's how inequality begins."

A similar idea was also mentioned in connection with **open research** by another participant who felt that making your repositories open allows people to have discussions about it. Open science is a movement to make publishing and sharing scientific research more accessible to all people in society. Openness was implicitly connected to other values like transparency, which is also about sharing decisions and data. To distinguish between valuing open science and other values participants mentioned, I only counted the value "open" when participants explicitly discussed the practice of publishing work in an accessible way or using open data in their research (i.e., when mentioned in connection to open science practices).

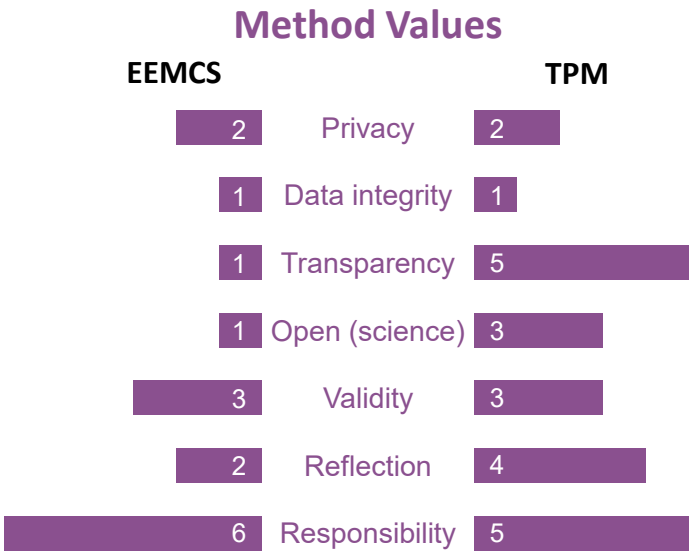


Figure 4.9: Method values: number of mentions per faculty. The number of participants from each faculty who mentioned a particular value in conversation, either implicitly or explicitly. A participant is counted only once, regardless of how many times they discussed a particular value.

Another value mentioned by participants was **validity**. Validity was described as making "realistic" or "reasonable" assumptions in your models and ensuring that "the conclusions you draw are actually sound conclusions." **Reflecting** on your assumptions, choices, biases and the purpose of your research was another process that people discussed as being important. One participant generally felt that "it's dangerous not to reflect enough" and a few professors shared that reflection is a practice they try to instill in students.

Responsibility was the most frequently mentioned method value (Figure 4.9), but not always in connection to the same things. For instance, three young professionals in EEMCS described

how they feel a responsibility to act in a certain way, as opposed to mentioning those actions as something they value. Others connected responsibility to a specific value. I summarize below five ways in which responsibility was discussed.

- **Responsible research practices:** Six participants talked about responsibility in terms of research practices. One person very generally discussed "the notion of best practice. So also, what to do, what not to do." More specifically, people described responsible research in data science as "checking the limitations of your approach," ensuring "we validate these tools in our data set or use case before reporting the results," and being "aware of what you are gonna solve for, what may be the possible impacts" if others misuse or misinterpret your methods or conclusions.

- **Responsibility to be knowledgeable:** Four participants discussed a form of responsibility that expanded on the idea of responsible research—that you have to have a certain amount of knowledge to be responsible in your research practices. One participant explained, "if you don't have full knowledge of the algorithms that use the data, you're kind of already in the wrong." A second participant was more blunt: "if you lack the base knowledge, then you are not responsible." This responsibility includes both being knowledgeable about data science methods and about "knowing the limits of the information you are creating."

- **Responsibility to reflect:** Two participants felt that data scientists, including themselves, have a responsibility to reflect on or question their research and assumptions in research: "It's our responsibility to question ourselves whether [our research is] really meaningful."

- **Responsibility to communicate:** Two participants mentioned that their research is uniquely positioned to bridge multiple disciplines. They both personally felt a responsibility to communicate their perspective in combining these disciplines with others in the community who may only see one side. A professor also mentioned they try to lead by example, which I interpreted as feeling a responsibility to demonstrate when they are in a teaching position—-another form of communication.

- **Societal responsibility:** Two participants specifically mentioned that data scientists (or researchers more generally) have a responsibility for our research to be meaningful or contribute to society.

Method values were the most shared between the people I talked with. Even though interpretations varied slightly (take for example responsibility), every distinct value related to the process of doing data science was discussed by at least two participants and by someone in both faculties (Figure 4.9). The values were also quite spread across different positions. Responsibility was perhaps the only value with enough mentions to consider variations between positions: it was mentioned by 6 out of 7 professionals, but only by 4 out of 11 students and PhDs. This seems to support the previously discussed idea that increasing expertise in data science is accompanied with an increased sense of responsibility in research (Figure 4.4).

Design Values

Participants named eight distinct values related to qualities that they felt were important to design into their research methods or results (Figure 4.10). Design values were not shared between the two faculties as much as method values were. Although I cannot say that the

different faculties don't value the same things, it does appear that they prioritize different qualities in their research outputs.

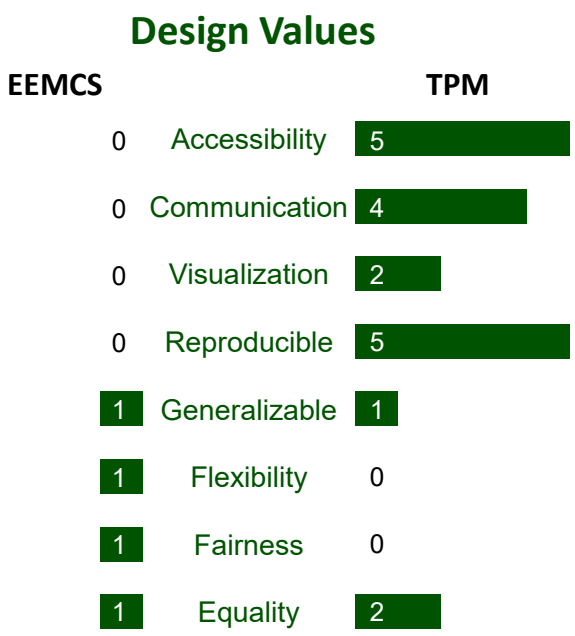


Figure 4.10: Design values: number of mentions per faculty. The number of participants from each faculty who mentioned a particular value in conversation, either implicitly or explicitly. A participant is counted only once, regardless of how many times they discussed a particular value.

Accessibility, **communication** and **visualization** were three values related to presenting research and data in a way that is understandable and open, and were only named by participants in TPM. One participant described accessible data as "not only sharing it, also sharing it in some way that is easy to use, easy to handle." Although similar to the value of open, accessibility also includes an aspect of preparing your data so that it is easy for others to use. Communication was a related value; one participant explained that "if the people cannot read [your research], then it's wasted." Beyond being explainable, the people who mentioned communication as a value also felt it was important to consider the audience of this communication. Specifically, they all also discussed that it is important to engage with stakeholders who are not data scientists—-with politicians, collaborators, professionals to whom you are providing models, or with general people or non-experts. Two of the people who valued communication specifically mentioned that visualizing models is an important (and valuable) way to communicate your insights.

Participants also valued making their work **reproducible** and **generalizable** in different contexts. One participant explained that making your work replicable can both ensure validity to your methods and results and adds meaning to the results themselves. Another person summarized this value by explaining, "if you do bad research with a good data management system, people can identify what was wrong and correct it." Two researchers also felt it was important to design their research to be generalizable: "I'm hoping that the new methodologies after validation in this domain could apply to different domains." Another related, but somewhat contradictory value described was **flexibility**. One professor explained that flexible algorithms are "developed in such a way that whoever uses them should know how to change their inner workings." This professor felt that "if you apply in different domains your algorithm, it's not the same algorithm. You have to change it somehow." Although related to the generalizability of algorithms, flexibility refers to designing methods that are readily changeable.

One participant also felt that it was important to make algorithms fair. They gave an example to explain that **fairness** is about designing algorithms so that the outputs are not biased in an unfair way. They also pointed out that "this topic has been around for a bit and it's still really heavily researched because it still needs a lot more research." A different participant critiqued

fairness (along with explainability and transparency) as a value, explaining that it "is a really narrow concept in computer science." However, it seems that the people I talked with gave more attention to other design values besides fairness. They also were able to give a broad interpretation of transparency – however, this interpretation did come mostly from people in TPM.

Instead of naming fairness as a design value, this participant felt it was important to consider "equality in terms of not only the outputs, or not only the access to the system or the environments to deploy the systems, but also in the production." **Equality** was implicitly mentioned in another instance, when a participant expressed that it is important to prevent inequality from happening. Fairness was also implicitly brought up when two people described competing against or being compared to others, rather than as an aspect of algorithms. Although these examples did not refer to fairness as a value (and are not counted in Figure 4.10), both participants expressed dissatisfaction with the lack of fairness. I elaborate on competition further in Section 4.7.

Value as Meaning

More than half of the people I talked with discussed that they believe it is important their research has **meaning for society** (Figure 4.11). As one participant explained, "the value is the impact" that your research can have. Some mentioned this explicitly, while others discussed this as a motivation for why they are doing research (see Section 4.4). Meaning for society was explained normatively – "our research needs to be practical" – and as a design value – "how do we make it usable for the community" or make our research "really like work for people's life."

It was mostly participants from TPM who felt that contributing to society was an important aspect of their research. This value was also mostly shared by PhDs and master's students (8 out of 10 mentioned this value, compared to 3 out of 7 professionals) – although this might be a result of talking to more people from these positions.

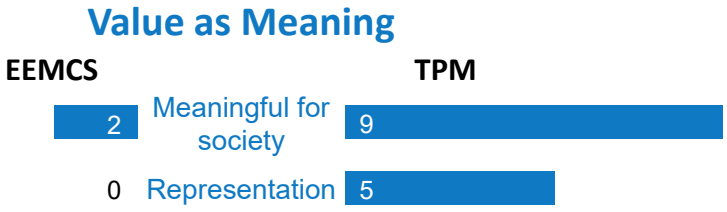


Figure 4.11: Values as meaning: number of mentions per faculty. The number of participants from each faculty who mentioned a particular value in conversation, either implicitly or explicitly. A participant is counted only once, regardless of how many times they discussed a particular value.

Representation was another value related to meaning in our research and work environment. Only one participant explicitly discussed representation as a value, connecting this to their identity as a person of color. They explained that "bringing someone to the table and giving someone a voice" is important because they didn't see people that looked like them in powerful positions when they started studying. They felt representation is an important value to make it easier for others to not be discouraged and to see themselves in certain roles. Implicitly, other participants also mentioned that they believe it is important that research represents more perspectives or areas of the world: for example, by prioritizing research in languages besides English or focusing on places in the global south.

4.3.2 Shared Values

After asking participants about their values, I followed up by asking whether they felt these values were shared by the rest of the community. I am using values as part of people's individual perspectives or worldviews. My intention with asking whether these values are shared in the community is to see if people feel like (some) of their perspectives are shared in the community.

Differing values between faculties

In Section 4.3.1 I present four figures that show how many participants mentioned different values, based on which faculty they are part of. The most clear differences between the two faculties were that more participants from TPM talked about design values and values referring to the meaning of their work (Figure 4.10 and Section 4.3.1). Participants from TPM on average also just discussed more values in our conversations, which may also have contributed to this result. At a glance, it seems like there are some values that are shared across all data scientists, and some that are shared more within a specific faculty.

A few participants also mentioned there is a divide between faculties or disciplinary groups. These divides were only drawn by people who were doing research in both TPM and another faculty. This could potentially be a result of these researchers having multiple faculties to compare, as more of the researchers I interviewed from TPM were working with multiple faculties compared to those in EEMCS. One PhD drew a divide between applied, engineering-focused individuals and the "mathematical and physics guys." They felt as though the mathematics and physics group did not place as much value on reflecting reality, practical applications, and validity. One PhD distinguished between the field of machine learning and the type of research happening at TPM, explaining that "machine learning looks only for correlations, not for causal chains that are more relevant for society." One participant explained that in considering inequality beyond fairness in algorithms, "especially computer scientists, they don't really pick up on this question... they probably just don't know what to do about it, so they don't really discuss it at all."

These last examples highlighted differences in values between the levels of research in data science (Figure 4.3). This difference is also apparent from looking at which values were most mentioned by researchers who are working on fundamental algorithms (Level 1), doing applied data science (Level 2), or are doing a bit of both (Figure 4.12). The three researchers working only on algorithm design primarily described method values; responsibility was shared by all of them. On the other hand, researchers doing applied data science were more concerned with the meaning of their work. They also shared method and design values, but these values were not as unanimously discussed. Those researchers doing both shared a combination of both values pertaining to the meaning of their work and about their methods.

Another difference between the faculties is how specific they were about naming which community they felt shared their values. It seems that people in TPM recognized that perhaps some of the values in their faculty (such as meaningful for society or representation) are primarily shared by others in their community, whereas people in EEMCS felt their values were shared more widely across TU Delft.

For example, when four people from EEMCS expressed that their values were shared overall, they did so very generally. One professor commented on the importance of responsibility in designing algorithms, "I mean, everybody thinks this is a big concern." A postdoc further commented, "I think nobody would disagree that you should make systems fair... Of course,

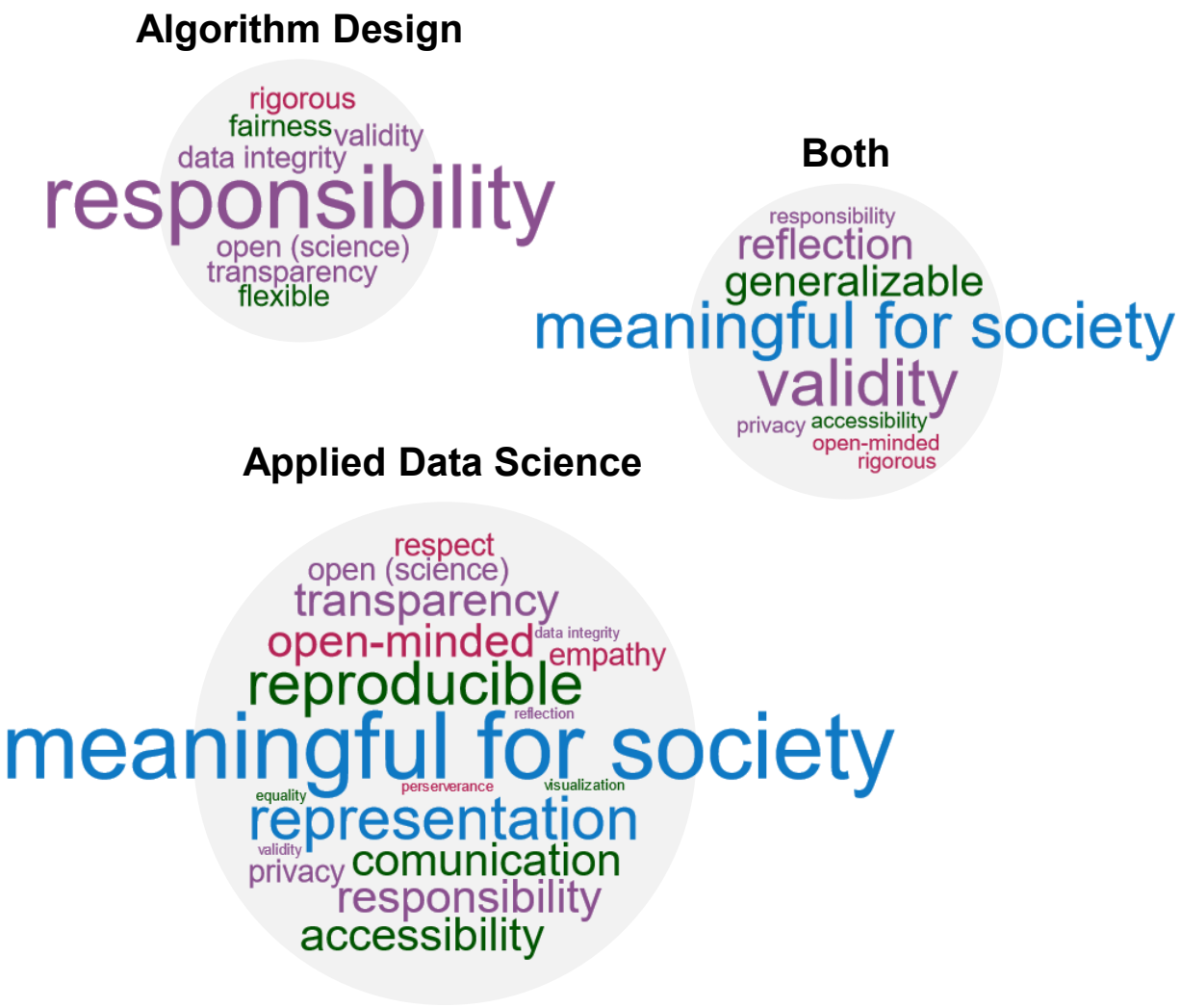


Figure 4.12: Most mentioned values by research type. Each word clouds show the most mentioned values by researchers working on algorithm design (3 people), doing applied data science (6 people), or both (3 people). Words are scaled based on how many researchers mentioned them, with the largest values in each word cloud shared by all researchers in that group. The colors correspond to the type of value: pink = personal values, purple = method values, green = design values, and blue = value as meaning.

I recognize there are differences across researchers, but I think overall in Delft itself, people tend to agree that it's an important topic." A master's student expressed that "the majority of people I've encountered share the same values as me." And a PhD explained that "the people around me and my colleagues also have many thinkings like me. We communicate with each other a lot and I think most of us, especially working in the data science domains, are very similar in thinkings" about responsibility and validity in their research. These individuals did not explicitly say who did and did not share their values beyond "people I've encountered" or people around them.

On the other hand, the people in TPM who expressed an overall feeling that their values were shared were also more upfront about who was part of their community that shared these values. One professor only commented, "I see that TPM is a very open faculty", whereas another professor explained that for transparency and reproducibility, "I see at least in my community, a

lot of other researchers are trying to do similar stuff. They have similar values." This professor explained that their community consisted of researchers working in a similar domain or on similar topics. Another master's student commented that responsibility towards society was generally shared, "at least in my EPA community." A PhD who is working on a project across two faculties also explained that ensuring validity and rigorous conclusions are values shared mainly in TPM: "I think that rigourosity started happening when I arrived to TPM. And I think that it's not the same, for example, because my project is [partly in another faculty]. But I think that it's really more enforced in TPM." This last quote is also an example of a PhD learning some values from their community.

Learned values

Three students from both faculties gave examples of how they learned about certain values in their courses. These courses solidified values that they already had from prior experience or education. One student who felt validity was an important value explained that their professors emphasized the need to be unbiased and take care when dealing with personal information. Two students explained that they were taught in their courses to be transparent about decisions, and that this reinforced their own values. One explained, "I've just documented everything that I did because that was something that was drilled into us," while the other explained, "if you wanna make a hypothesis, just prove me if your hypothesis is right or wrong. Just being transparent, that's what I get from TU Delft." Both validity and transparency were method values mentioned by a third of participants.

PhDs from both faculties also on occasion described values as learned from or shared by supervisors. One PhD explained that they learned the importance of validity from their promoter: "actually, it's [my promoter] who made me think I cannot just do some research to make the statistics and then draw the conclusion." Another PhD felt as though their supervisors have an even greater sense of responsibility to do contribute to society than they do: "I feel like they have the responsibility for the whole world."

There was only one instance of the reverse, where a postdoc described influencing a professor to think more about a value, "now [the professor] was researching it a bit because I was working on it and [the professor] was working with me, but otherwise, [they] wouldn't have gone towards this kind of research directions."

Professors have an influence on the perception of shared values, especially through teaching and supervising. They are able to reinforce certain values in the community, and that is felt by students and PhDs especially. Between professors, there is also an overall feeling at TU Delft that their values are shared by their colleagues.

Differing values between levels of expertise

There were only two examples from all of the interviews in which a professor/supervisor did not demonstrate the same values as the supervised student. However, these examples were not explicitly mentioned in connection to shared values. One was mentioned in connection to support in relationships. In this example, the student worked on a project that they felt was meaningful for society and connected to their value of equality. The feedback they received from individual professors was that "you should have never worked on it, like it's too political." This student felt that this "might be the kind of attitude the faculty [EEMCS] is promoting, that you shouldn't do things that are considered political or activism." The second instance was

mentioned in connection to identity. This student, a person of color, felt it was important to highlight cultural and social differences in their research, even when their supervisor encouraged them to change their scope to have better data. While no values were explicitly mentioned in this story, the student explained that "when I look at a lot of research, they only control for income, but they don't look at how this may affect people of color and stuff like that. So, I've made it a priority to deal with stuff like that as well." I can infer that their supervisor, however, prioritized having available data and ease of processing over representation.

Although professors were not singled out as a group that did not share certain values, students were. This is part of how participants distinguished between levels of expertise in the data science community (see Section 4.2.1). One professor explained that "you try to teach these things that you should be responsible for your problem. But quite often I feel some students might just want to solve the problem. By any means. And this is the only way how we can explain these values, is if we state them as a regulation." This professor felt that when you get to PhD level and above, people start to feel more ownership over their research and pay more attention to responsibility and validity. This was supported by a postdoc who felt that their values of communication and visualization were most shared by people with a certain amount of knowledge: "individuals who have a high caliber in understanding are the ones that really appreciate that we've got this data, let's work with it and see what it tells us." A master's student was more direct, explaining that data science "methods and tools that are so powerful and so commonly misused that, especially by master's students in their thesis, I would say, which is for me somehow problematic." There was also a PhD student who pointed out that they struggle to live up to their values of reproducibility and accessibility: "even myself, you know, like I have these values. But I'm for sure not educated enough to properly write my code and data in the way that is understandable." This PhD felt that overall, these values were not shared, or at least not prioritized enough.

Differing values between institutions

Industry was also highlighted as having different values than academia. This difference was highlighted by young professionals in both faculties. "For industries, they want to use the data and make money because that's the nature of companies. But the universities, we are concerned less about money." This young professional felt as though industries did not share values related to meaning in research. Another participant felt industry did not place enough emphasis on data privacy. Other participants felt that industry especially does not share the same responsibility as academic researchers: "industry? I don't feel people have this responsibility so much. Many of them only care about the money they could get from the project, but do not care about the quality of the work they can produce."

There was one young professional who felt the opposite, that industry might share their value of fairness more than academia does: "people realize how important it is. And some companies might just do it because they feel like they have to, but I feel especially in research also, this is a big topic and a lot of people do research on it." Based on how fairness was critiqued by a different participant (and the fact it was only mentioned by this young professional), their impression seems accurate.

There were also two young professionals who also expressed that TU Delft as an institution might not share their values. One felt TU Delft does not demonstrate a responsibility to be knowledgeable (by ensuring its students are knowledgeable), "because I've seen the sheer numbers of people that TU Delft wants to graduate." A different young professional also felt

that TU Delft did not prioritize data privacy enough: "It's just not taught enough at school how important it is. I mean, these days it's getting better." This person also felt that the European Union also did not prioritize data privacy enough. Contrary to this person's feelings, a different student felt that the university "is really explicit [about] the data consent forms and privacy thing." This student described how data privacy was a learned value from their supervisor, because they kept being reminded about it. Another student similarly felt that professors, "when I'm doing my internship, they really put much, much emphasis on stripping data from personal information."

4.3.3 Values: Discussion

Based on my conversations, there are a lot of similar values circling around the community, but they are perhaps not all prioritized in the same way. When people felt as though their values were not shared, it was not so much because they thought others disagreed with that value, but rather that they don't have the same priorities. This was especially evident in differences between what researchers working on fundamental algorithms valued compared to researchers doing applied data science. It makes sense that people doing applied research are prioritizing the societal impact of their research, while people who are developing (potential) methods or tools are concerned more with *how* they are developing these tools. This difference is not a bad thing; in fact, it emphasizes that there is a diversity of perspectives in the community.

Computer scientists were singled out as having a different set of values, mostly by people from TPM or who were involved with multiple faculties. This appears to be an accurate assessment. People from EEMCS did not discuss many of the values mentioned by people affiliated with TPM, named less values overall, and also had less shared values between them. Except for responsibility and validity, the values they did talk about were discussed by at most two people. This lack of consistency could be an indication that there is less conversation happening in EEMCS about values and what they mean. As one participant explained, "we don't really tend to name values."

Take for instance the fact that there was greater consensus about method values. Perhaps this is because conversations about what is important in the process of doing research are normalized in the community already. However, it also highlights that we could benefit from talking more about personal values, design values, and the implications of our research. Thatcher et al. (2016) explains that when big data is critiqued, the critiques usually focus on limitations of methods or failures to produce results. By neglecting more socially-oriented critiques, or in this case more socially-oriented values, we might also neglect the influence of our identity and environment in shaping research. Our workplace can influence our ability to do research, so it is also important that we can name and talk about values like respect and empathy.

Naming values is an important part of expressing our worldview in community. It is one way to ensure greater representation. Sharing our perspectives can demonstrate to others the relevance of these perspectives for the overall community. This is being done by professors in TPM, who are teaching certain values to their students. Professors have an ability to shape (or confirm) student values, but there is less of the reverse happening. As Kirkness and Barnhardt (2001) highlight, students are adapting to the system more than the academia is adapting to become relevant to them. This suggests an actionable way to address representation of students in the community: professors can actively listen to and repeat back the values their students express as a way acknowledge their importance.

More open discussion can also help with understanding interpretations and the context in which values are used. Many of the things valued by this community are aligned with an inclusive and decolonial perspective of research, such as taking responsibility of the impacts you have and sharing knowledge by making your results accessible and open. However, to actually act upon these values in a decolonial way, we have to take the time to align what we mean by these terms. This is especially important to counteract the "narrow" interpretations of values like fairness, explainability and transparency. An example of how narrowly defined values can be problematic is explained by Gasparotto (2016). They explain that although transparency is an important guiding value in most literature on AI ethics, open-access knowledge can be interpreted as problematic from a Native perspective because it is seen as extracting valuable knowledge from their communities without compensation or control (Gasparotto, 2016). Although this is a negative example, expanding the definition of values can also lead to positive opportunities. For instance, the many definitions for responsibiliity that participants provided can make an this abstract value more actionable. To be a responsible researcher, you need to 1) learn your subject thoroughly, 2) reflect on the limitations and contributions of your research, 3) communicate your research to others, especially those outside of data science or your discipline, and 4) consider how your research can provide meaning to society.

4.4 Agency, Identity and Reflection

In this section, I look explore the third sub-question I ask in my research – how identity influences people's ability to express their perspectives in community. This question is about representation (through agency), identity, and recognition (through reflection). Although I asked about these three concepts in separate questions during interviews, I discuss them together because the concepts are also intertwined.

4.4.1 Agency

Agency, as defined in Chapter 2, is about the ability to act in ways aligned with our values and to shape how our perspectives are represented in community. Asking about agency to express perspectives is an abstract question. To make it a bit more concrete for participants, I asked, "why did you chose to study or do research in data science (or a particular area of data science)?" This question revealed something about their motivations and whether there were other pressures pushing them into a certain direction. Being able to act upon our motivations is a form of agency. This was also one of the first questions I asked participants because it allowed me to learn more about what they are doing as well.

Curiosity was the primary motivation people mentioned, which highlights that overall people are able to pursue research that interests them. However, many other factors were also mentioned, ranging from economic incentives to the influence of individual people. I will mention the external factors first, then return to curiosity and various internal motivations.

External Motivators

Money: There were three people who talked about financial benefits influencing their decisions to go into computer science. All three of these participants were international (non-Dutch), and two explicitly mentioned that in their countries, there is cultural pressure to go into more technical fields. One explained that in high school, if "you don't really know what to do but you have good grades, you're told to go into a more technical track... then you can more easily find a job." The other explained that they grew up around the rhetoric, "if you're not a doctor, lawyer or engineer/ITist, then... you'll be starving to death." These two participants were raised in environments where data science was highly valued as an economic opportunity. A third participant commented that they entered data science because they personally saw an economic opportunity: "I think in the new age, data is the new oil. So it is very valuable."

Opportunity: Opportunity – not only economic but also life opportunity – was another motivator influencing people's decision to pursue data science. Participants talked about going into research or the position they are in because an opportunity presented itself, or because it provided more opportunities for their future. Job prospects or the potential to pursue a certain career were also a factor affecting people's research focus. For example, one student decided to pick a specific track in their program because "I also wanted to keep my options open." Another PhD who pivoted from a different discipline to data science explained, "I saw the job opportunity here, that they are hiring people to study [topic]. I think that's a good opportunity to expand my research." This sentiment was shared by at least two others – a position presented itself and the participants took it.

Taking these opportunities was a way that these participants ensured greater agency in their research. As the student explained, pursuing data science is a way for them to branch in many

different directions in the future. A professor similarly explained that the position they are in now provided the opportunity to explore research lines that interested them, "I moved more to a higher level where I have more freedom."

Necessity: On the flip side, two students mentioned that they ended up in data science somewhat out of necessity, or because of a *lack* of opportunity. Both master's students in TPM, one mentioned that they started exploring data science because it was required content in their studies. The other discussed that for their thesis topic, "I was abroad and I kind of forgot that you need to start really early with topic choosing. And because I was abroad, I couldn't go to the events to find something." So they took what was available to them, which happened to be through a professor that hadn't posted a topic yet and offered it to this student. The lack of agency in this second example came from *not* having the opportunity to join an event to select topics; because of a lack of access, this student also had less choice.

Influence of others: A few people also mentioned that specific people influenced their research path. One professor explained, "I was wondering what I should do with my life, and my advisor told me, if you want to do a PhD, you can do." A student picked their thesis topic "just because of a professor. I met my current supervisor and [they] were extremely passionate about what [they are] doing." For both of these participants, people who were supervising them encouraged them to continue (or start) research in the area they are currently in. This also suggests that supervisory roles can have an influence over a person's agency in research.

There was one student who also mentioned that their peers partially influenced them. Their curiosity about their research topic was piqued because students around them were hyped about it.

Curiosity and Internal Motivations

Having the freedom to choose comes with flexibility. When I asked a professor why they stay in academia, they responded, "I love teaching. I love pursuing my own interests." Curiosity and interest was the primary reason people purused a certain line of research (or study). It was mentioned in 12 of my 18 interviews. For most, they were curious about exploring different methods in data science. PhDs and professors from both faculties shared this sentiment, explaining: "I feel like machine learning and deep learning technologies are really interesting;" or "I got pulled into more of the method sides because I found them so intriguing."

Beyond just interest in the methods, participants also expressed an interest in understanding phenomena in the real world. For some, their experiences directly led them to research specific topics in data science or to learn more about data science methods. This was expressed by young professionals and professionals in both faculties.

Some participants expressed that they had a general curiosity to understand the world. One participant got into data science because was curious about complex systems and understanding nature, while another found in applied data science, "I saw the human side of it, which I didn't see for other disciplines."

For three people, their research motivation was directly connected to problems they saw working in industry or government that they wanted to solve. One professor explained, "I thought that management could be done much better. And then this kind of sent me into this route where I discovered methodologies and techniques that were very flexible and very powerful to tackle planning and management problems." This quote also demonstrated this participant's desire to

have an impact on society.

Societal impact: Having societal impact was not only a value, but also a motivation for many participants. A professor summarized a sentiment shared by multiple participants: "I wanted to help with something, this problem... because you know, these problems actually affect our society." Through understanding problems, these participants hoped to provide insights that would be useful.

This motivation was demonstrated by one participant who explained why they moved from a study related to climate change to a study more focused on data science and policy in TPM. They felt that in climate research, there was a mismatch between research and policy. Their current research, however, "might raise more interest to policymakers, local authorities, and so on." This participant values having an impact on society and made a choice to switch studies because it aligned better with this value. This was a clear demonstration of agency.

A desire to be more interdisciplinary: Quite often, participants expressed that they were interested in merging interests or bridging a gap. Many of them felt that data science provided a way for them to bring together these interests. This feeling was shared by all participants, even learners. Two students who had not yet started research of their own still recognized the potential to combine data science with other disciplines. It was a reason both of them decidd to pursue data science: "when I first started doing computer science, I wanted to combine somehow biology with computer science or any other disciplines... because it could ease up the process of everything with computational power."

Eight of the young professionals and professionals I talked with were actively combining disciplines (which I attempted to visualize previously in Figure 4.7); I would consider their research to be either inter or transdisciplinary. For some, this combination was about interest. One person moved from computer science to applied data science because they were "not extremely excited, and that's why I started to think more about other disciplines and combining disciplines." Another chose their research topic "because it combines the two spheres. So I think it was the most interesting thing for me."

Merging disciplines was also for some participants about merging their abilities. One participant explained, "I saw myself as this interesting bridge between methodological researchers who only care about making things scalable, efficient and so forth but the have no interest in real world applications, and... the water systems [part, who] are very constrained with their tools." Another young professional explained they chose a PhD that allowed them to combine their background in both mathematics and machine learning. Many young professionals especially talked about moving from one discipline to data science, and wanting to carry their previous background with them. Data science appears to be a field in which it is easy to combine disciplines. One participant explained of their own transdisciplinary research, "data science is a bridge to link the two domains together."

However, as I mentioned in Section 4.2.2, having a community that intersects many disciplines can also be challenging. Although people in this community have the agency to combine their interests, this was also seen as a challenge by one professional. They explained, "I consider my research interdisciplinary, so I don't conform to one research line. That also makes it very difficult for me." For them, doing interdisciplinary research was also part of their identity, which impacted how they navigated a research world that is structured around disciplines.

4.4.2 Identity

Identity – whether that is social identity or our experiences – shapes our worldview and perspectives (Chin et al., 2022). I had a two-fold purpose for asking about identity in relation to data science. First, I wanted to understand how identities might influence the ways people approach their work as data scientists and the perspectives they bring to the community. Secondly, I wanted to understand what identities participants bring into the community.

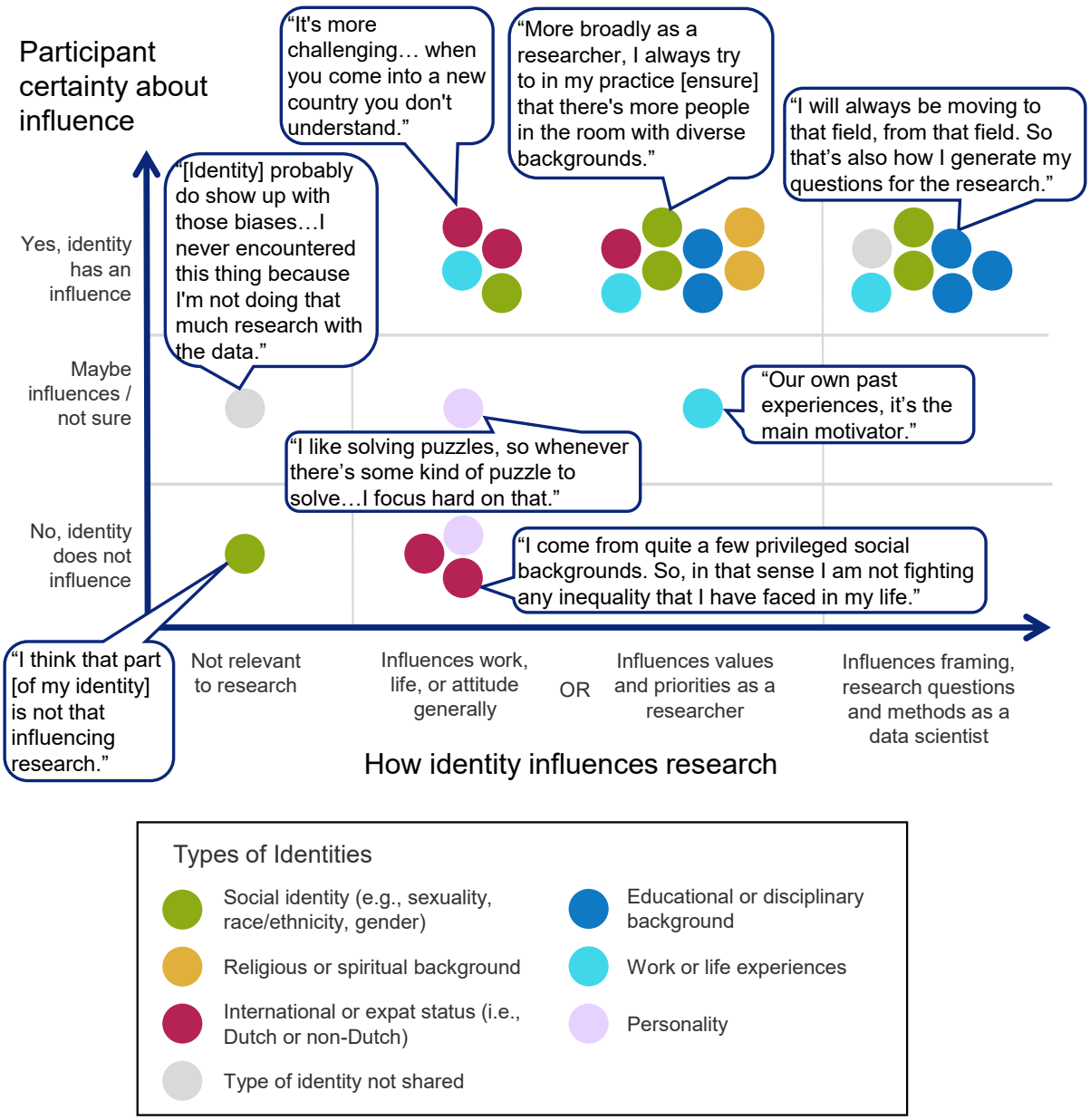


Figure 4.13: Influence of identity on research. This spectrum shows how participants answered the question, "what aspects of your identity impact how you approach data science?" The color of each circle corresponds to different types of identities that participants mentioned. Participants who mentioned multiple identities are represented by multiple circles. The horizontal axis shows a few generalized ways that participants described how identity influences their research. The vertical axis indicates how certain participants were about whether their identity influences their work at all.

In each interview, I asked the open-ended question, "what aspects of your identity impact how you approach data science?" Similar to the question about values, a few participants asked

me what I meant by identity. I took identity in the broadest sense possible, explaining that it could include anything ranging from gender to political beliefs to educational background or life experiences. Although I was curious about what types of identities people would bring up, most of the responses focused more on how identity impacts their research, rather than which identities impact it. Two participants actually pointed out that the framing of my project already primed them to answer this question about identity in a particular way. "If I'm really grasping what you are trying to ask, yeah. Having an other sexual orientation is what makes me different, apart from the others." This person went on to discuss that possessing a minority identity changes how they approach work and life in general, rather than data science research specifically.

This was one way that the influence of identity was described by participants, that it either impacts their attitudes to work or life generally or broadly influences their values and motivations as a researcher. Other people were more specific and described how identity directly impacts their framing or scoping of research questions or the methods they take as data scientists. There were also a few who felt identity was not relevant to their research at all. However, there was also a second aspect to how people answered this question, and that is how certain they felt about whether identity influences their work or research at all. The responses to this question are summarized in Figure 4.13. In the following section, I explore this spectrum in more detail and look at the different types of identities that came up in conversation.

Identity does not influence approaches to data science

One of my assumptions is that identity has an influence on how people approach data science. There were three participants whose initial response suggested they disagreed with this assumption. Despite their initial feeling that identity does "not really" or "not particularly" influence their work, all three still ended up describing a way that identity does influence their research. These responses correspond to the bottom half of Figure 4.13, where participant certainty was "maybe influences/not sure" or "does not influence."

Two of these participants connected identity back to their motivation for pursuing a particular line of research. One participant explained that the influence of identity is not something they had thought about before our conversation. "If my identity comes up in my research, I would maybe think about things like what I said before, like bridging a gap. I think something is cool and... maybe fits my personality a little bit." Another participant similarly responded, "I don't know, my identity? Look, how do I say that—so I always say I like solving puzzles... then I focus hard on that." Although initially unsure how to answer, both of these participants were able to connect their personality to their desire to solve problems and understand fundamentally how things work in their research.

The third participant had a similar initial response. However, they quickly revisited their answer, reflecting on how they have "quite a few privileged social backgrounds... and then my private life doesn't have a negative impact on my working life." This participant recognized that identity may impact their experience working, but felt that it doesn't convey "advantage or disadvantages in terms of research." This feeling was echoed in a different conversation when I asked a participant if they identified as a person of color. They explained that although they feel different living in the Netherlands "which is a white people world... that part is not that influencing research."

The initial responses of these participants may have had more to do with uncertainty around

how to answer the question, rather than actually disagreeing with my initial assumption. For example, there were two other instances where the participants expressed confusion, responding that the question made them feel a bit "lost." In both of these conversations, they were unable to provide an answer based in their own identities, but still acknowledged that identity could have an influence over research. One student answered that identities "probably do show up in those biases," but felt the question was not really relevant for them, since they were not doing research. The other participant gave a general example of a harm happening in society that was "the main motivator" for them to pursue their research subject. This person expressed that they could not think of a more personal experience, yet they still made a connection between real life phenomena and how that influenced their choice of research direction.

Identity impacts framing of research methods and questions

On the opposite corner of the spectrum, some participants clearly connected aspects of their identity to how they choose or frame research questions and methods (Figure 4.13). Three participants brought up educational background, one discussed work experience, and two brought up social identities that they had a desire to represent through their research questions and hypotheses.

One participant explained the influence of identity on forming research questions clearly: "how we can construct hypotheses is different based on and influenced highly from their perspective, their personality." This person shared they personally have a genuine interest to ask questions about how women interact with society. A second participant described that focusing on people behind the data, especially social and cultural differences was important to them and something they prioritized in their research. "I think it's important to represent, because there's such a bias towards like global north and Western countries because it's easier, because there's better data available." This person explained their conviction to represent came from their background and "the fact that I kind of live between cultures." In both of these cases, the participants expressed a desire to ask questions about a social identity that they possess.

Educational background and work experience were also mentioned as aspects of identity that influence research questions and priorities. One participant explained "I always be close to transportation... I think that always I will be moving to that field, from that field. So that's also how I generate my questions for the research." A second participant chose to work in data security because "I've seen the data are not being protected properly." They explained this research focus is primarily from their experience working in industry.

Beyond just shaping research questions, educational background was described by three people as influencing how they frame or approach their research. One participant explained that "my background is majorly engineering and there are some people in this department or in this whole field, they majored in mathematics and physics. This makes two different approaches of research." Another participant described that being educated "in a really strict way academically speaking," impacted how they frame knowledge and share what they do and do not know. For the third participant, they explained how their background in computer science made them feel under-prepared to use different methods in their current research. "I feel like I've never been trained for this kind of methodology and for reflexivity and all. So I don't think I'm really good at it for now." They described how identity both influenced their ability to use certain methods in research, and that it solidified their value for reflecting on their own biases.

Identity shapes values and perspectives towards work

Many of the participants described how identity has a general impact on their values or perspectives. As one participant explained, "I'm not sure if [identity influences me] as a data scientist particularly, but more broadly as a researcher." In these conversations, people mentioned upbringing, gender, sexuality, and international status as aspects of identity that impact their work broadly.

The experience of living in different countries was brought up by both Dutch and non-Dutch participants. These participants discussed how international status can influence both ease of living and attitudes towards life. One of the participants who was initially unsure about the influence of identity over research explained that being half Dutch makes it "much easier to live in the Netherlands." Another participant, reflecting on communications with their collaborators, observed, "in a lot of ways, I'm pretty Dutch in the sense that I tend to be direct. And what I notice is with colleagues that are Dutch that it makes a lot of sense, but with our colleagues that are non-Dutch, I don't know. Maybe they have get used to it."

These responses were from people who were uncertain about the influence of identity, whereas the people who were certain about the influence of international status could all be considered expats (FIGURE X). One participant explained that being an expat comes with both perks and challenges, because you are in "a new country you don't understand, where you are trying to... understand the place and how it all works." Living in multiple countries shaped their attitude towards facing both challenges in life and in research: "you get on with it and... you just have to learn." Similarly, another participant described how living in two countries prior to coming to Europe was "more motivation for me to be open-minded to the whole world." Being open-minded and persevering through challenges were values these researchers held.

Two professionals also shared about how upbringing shaped the values and personality they bring into work. "I think in general as a professor you cannot split your persona from your teaching persona." This person explained the respect they bring into their work comes from their religious upbringing and previous education. Similarly, a second participant reflected on how their sense of self came from their spiritual grounding and having good role models, family, and community. "I think if I didn't have a strong identity with my own self, then research wouldn't be for me."

In the examples above, participants connected identity to personal values and working conditions. There were a few instances where participants mentioned having a social identity that made them feel like a minority or outlier. These participants talked about how identity shaped their perspective and how they see the world, both in research and in life. Everyone who mentioned a minority identity (including the person who felt being a POC does not impact research) was certain that some parts of their identity influences their work, evident by the presence of these circles at the top row of Figure 4.13.

One participant explained that they identify as gay, which "makes me different, apart from the others" in the data science community. They explained that feeling different socially makes them consider "differences and the things that we have in common with others. So basically, that's what it is with the data also." They later clarified that identity impacts not data science necessarily, but how they approach work more generally. Another participant who identified as a woman in computer science and an international at TU Delft clearly explained this connection between being a minority and having a different perspective:

> "There is causality here. If you are a person with a diverse background, you're more likely to be woke about certain things. The more woke you are, the more you think about certain aspects of your work. So of course I get that the more woke I became with my research, with life in general, the more I started noticing things that really bother me in the research of others or behavior of the others as researchers."

This participant went on to explain that they then have to decide whether or not to confront the behavior of other researchers. This was not the only conversation where we discussed examples of confronting others or standing your ground. Identity affects more than just perspectives; it also impacts some people's sense of responsibility to speak up about or represent a certain perspective.

Responsibility to represent

One form of responsibility discussed by participants was the responsibility to communicate a certain idea or perspective (Section 4.3.1). The researchers who shared this responsibility found themselves at an intersection of multiple disciplines, and thus possessing a unique perspective to share. A similar feeling was also expressed by a few participants who identified as part of a minority at TU Delft.

One young professional who identified as multicultural and a person of color explained, "I feel like it's my responsibility to remind people like, there's just this full other dimension of things that should be included in research." They described an occasion when their perspective towards their research lead to a misalignment in priorities with their supervisors. "I like kind of stood my ground and was like no, I really want to do this because I think it's important."

I asked another participant about what circumstances prompt them to confront or not confront the behaviors of other researchers. They answered that they are more likely to speak out when it is clear the researcher "just never thought about it," or the behavior "affects minorities I belong to." They explained that they are more conflicted about whether to say something when it is to "more senior people," when "somebody does something repeatedly," or when "they don't agree with you that this is important."

A professional discussed the importance of using their position to ensure representation as well, not just to speak out as an individual. They explained that they try to "practice that there's more people in the room with diverse backgrounds... So, whenever I'm in a situation where I can hire someone, I'm bringing someone to the table and giving someone a voice." They explained that they felt like a weirdo during their studies, often being questioned because they are a person of color and a woman. Now they want to make sure that other people have it easier and "can see themselves in certain roles."

4.4.3 Opportunities for Reflection

My interviews were opportunities for participants to reflect on the values and identities they bring into data science. I was curious to learn about in what other contexts people in the community reflect on these questions or on other assumptions and biases they might have. This was also a way to learn about whether there is recognition within the community.

Most people described reflecting in informal conversations with peers. There were less mentions of structured reflection opportunities such as courses for bachelor's and master's students.

Structured Reflection Opportunities

Courses were mentioned by students and professors as a place to practice reflexivity. Two students from EEMCS recalled reflecting on biases and assumptions in a course on ethics in machine learning. However, they felt that outside of this course, these conversations about biases didn't happen between students. "We students between each other, we didn't discuss so many things about biases. We discussed it when it was required. . . but otherwise not really." One of these students further commented "I don't know if we need more" reflection opportunities.

While students from TPM did not mention specific courses, one student felt that there generally were opportunities to reflect. Two professors in TPM also talked about how they emphasize reflexivity in their own teaching. One professor explained that "thinking about those choices and the implications and the consequences is what I really like to instill in students." These professors are creating opportunities in their classrooms for reflection.

Outside of classrooms, the only mention of a structured reflection opportunity was from one young professional who attended a few workshops in TPM or IDE, where "I really noticed people were talking more about it."

Informal Reflection Opportunities

Most of the reflection opportunities described by community members took the form of informal conversations. Six people mentioned they reflect with their peers or collaborators. The peers or collaborators described were typically those in the same position: a professor described reflecting with "with my collaborators mainly," a master's student reflected with "my community in EPA," while a PhD candidate reflected regularly with "a really good friend sitting next to me in my office" who started their PhD at a similar time. One student was even more specific about who they reflect with. Identity is something they only discuss with "my women friends. . . and also with my gay friends," the latter being a group with which they share not only position, but also a social identity.

This is in contrast to only three young professionals who mentioned reflecting with someone in a supervisory role. All three mentioned that their supervisors in TPM or civil engineering played a role in helping them understand their own biases better. One explained that "actually it's [my promoter] that made me think I cannot just do some research to make the statistics and then draw the conclusion." Another participant explained that their supervisor "also helped me reflect about my own bias as a computer scientist."

There were even less mentions of people reflecting with those in a lower position. One postdoc explained that they are trying to create more opportunities for professionals to reflect on their own research. This participant felt as though they had the capability to provide these opportunities for computer science professionals especially, because "we share more or less the same vocabulary, I know what they are working on, so I can try to also develop maybe frameworks for them to reflect on this." This was in contrast to a master's student who talked about challenging the biases of their colleagues, both those who are in a similar position and young professionals who are further in their careers. When I asked if they feel they can create reflection opportunities, they responded: "I don't think I can create opportunities. I can only create them for myself and that's because of my position."

Limited Opportunities to Reflect

There were three participants who felt there are not many opportunities to reflect at TU Delft – all three affiliated with EEMCS in different positions. This was not the experience of all researchers in EEMCS, as there was also one participant who shared that they reflect on their research almost every day with colleagues. However, especially for young professionals and professionals in EEMCS, there was only mention of informal opportunities.

Two participants did not comment on whether there were opportunities present, but rather expressed they "haven't had the opportunity" to reflect with others at TU Delft. One was a learner, and the other was a professional who is in a new position.

Not everyone viewed the importance of reflecting the same. While almost a third of participants implicitly or explicitly discussed reflexivity as a value, there was one person who felt that reflecting on assumptions in their research was something "I haven't really had the need" to do. I asked if their collaborators doing different research would come to them to reflect, and they similarly shared that others don't typically come to them to discuss these topics.

4.4.4 Agency, Identity and Reflection: Discussion

Reflection opportunities exist in both faculties. Unfortunately, I did not get to talk with any professors who may be guiding the kinds of ethics courses mentioned by EEMCS students. However, the impression I got from participants in EEMCS is that reflecting is not as big of a priority as it is in TPM. This finding might also be an unintended consequence of how I asked about reflection opportunities; sometimes I asked about reflecting on identity, other times I asked about reflecting on biases and assumptions. In my mind, these are all interconnected and fall under the same umbrella of reflection. However, for those participants who feel as though identity is less relevant to their work, the importance they place on reflecting might change depending on what they are reflecting on. Reflection, like many of the concepts I asked about, can have multiple interpretations. The significance of reflexivity is something that I take for granted; it is emphasized heavily in my studies and evident in how I write. However, I cannot assume that participants understand reflection in the same way I do.

Yet, this points to the way our environments can shape us. The professors in TPM actively emphasize reflexivity in their classes; in doing so, they also influence how others (especially students) value reflection. People who described reflecting with their supervisors also shared that they learned something in that process. In my framework, recognition through reflection is a prerequisite for meaningful inclusion. Creating more opportunities to reflect can help the community see the value of reflecting and lead to conditions that support greater inclusion.

From my conversations, reflection opportunities tend to be initiated by professionals. There's an opportunity here to empower those in lower positions — the learners and the young professionals — to feel as though they can create reflection opportunities, drive conversations, and call people out. This might lead to more far-reaching impacts than imposing formal reflection opportunities in courses or workshops. Most informal reflection opportunities are with those in a similar position. If my own experience is any comparison, I also tend to reflect on my work primarily with peers. If we focus more on encouraging students to reflect, especially with their peers, then perhaps more of the community will engage in the recognition necessary for participation.

Feeling as though we can create opportunities is related to agency and feeling like we have the skills necessary to act in ways aligned with our beliefs. One participant explicitly told me

they did not feel trained in the skills to reflect, despite believing data scientists need to reflect more. However, when it came to research, participants described having a lot more agency. The data science community seems to be a space where researchers have a lot of room to express themselves. Participants I talked with are pursuing things that interest them, are doing research they find meaningful, and are able to bridge disciplines in ways that acknowledge societal complexities.

For a few participants, their identity played a role in what interested them and how they shaped their research questions. This connection highlights the importance of incorporating diverse perspectives in research communities. People go into research based on their identity, their experiences, and how they understand the world. Literature also points to ways research underrepresents certain regions of the world or groups of people. Franklin et al. (2023) found that the journal *Geographical Analysis* lacked research explicitly about women, despite the fact that so many aspects of geographical research are gendered or concerned with social inequities. By bringing in more representation, our research questions will likewise expand the ways we can learn about society and the types of meaning we produce. But we can also consider the reverse: by focusing our research questions more on subjects, peoples, or systems that are too often overlooked, perhaps we can create more space for people underrepresented in academia to find a place they belong.

Participants often demonstrated their values through the way they described their research. Many participants with an applied element to their project were motivated to do research so that they could contribute to society in a meaningful way. Two people doing fundamental research were motivated to solve actual problems they saw while working in industry. Instead of describing their motivation as meaningful for society, they were driven by a desire to address difficulties they experienced in industry. This is aligned with the value of responsibility, especially described as a form of best practice.

Being able to bridge multiple disciplines was a very common reason why people chose to pursue their line of research. Data science provided an opportunity for them to work interdisicplinarily; however it also posed a challenge. Some researchers who classify their work as interdisciplinary feel less understood because they don't fit into the disciplinary structure of academia (I provide more results to support this idea in Section 4.5). Perhaps creating more cross-disciplinary opportunities is a ways we can make it easier for people to exist between and across disciplines. The structure of the university itself (having separate faculties) does not reflect the way people incorporate many different disciplines – the university is not structured to support differing disciplinary perspectives in the same spaces. To create more cross-disciplinary opportunities is a way of validating all these researchers who do bridge multiple interests, domains, or disciplines. As with reflection opportunities, this action might also reinforce the value of working with diverse disciplines or perspectives. This builds on how one participant described the importance of representation: having representation in different spaces can make it easier for people to find others who are going through similar challenges. Returning to a disciplinary representation of data science, it seems that people form the intersections between various micro-communities within the umbrella of data science.

Almost all participants felt at least some agency in their research. However, there is still external pressure that shapes the actions we can take. In some instances, these pressures were connected to identity. Two international students described being encouraged into STEM fields, specifically computer science. In one of my conversations, a professor mentioned that "students from the Netherlands perform poorly compared to students from abroad" because in

other countries they require a much higher level of education prior to university. I do not have any reference to compare the level of education between different European countries. But based on the comments from these students, perhaps the higher level of education is a product of societal cultures that consider STEM fields and computer science as superior professions. This type of societal pressure is growing and shaping the data science community here, even if it is not felt by all of the data scientists I talked to. Most of the researchers see the opportunities in this field and some even moved into this field because of the opportunities. This is an example at a societal level of how providing opportunities can also shape what we value.

Assuming that opportunities can shape culture the same way that culture can shape opportunities makes changing culture more actionable. Acting with agency and representing your story is a way individuals can create a more inclusive culture. This is demonstrated most clearly in how participants from minority identities described feeling a responsibility to speak up for what they believe in. These people have an incredible strength. It made me wonder, how might we build confidence in others to realize what they can contribute, what parts of themselves they can also represent?

4.5 Relationships

I place an emphasis on relationships in both my framework and sub questions. My conversations further supported the idea that relationships play an important role in how we navigate community. In this section, I expand on ideas I presented in Section 4.2.2 by exploring what makes certain people "our people" and closer to us in a relational representation of the community.

I did not prepare any questions about relationships for my interviews; instead, I looked for opportunities during the conversation to learn more about the relationships participants mentioned on their own. I often used the drawing activity to probe participants about what made certain relationships feel supportive or what made participants feel well-connected to others. Some mentioned having "their people" that they felt close to, supported by, intimate with, or with whom they could share about either research or life. What surprised me most is that many participants had very similar answers about what made them feel close to or supported by others.

4.5.1 Supportive Relationships

In some of my interviews, I asked participants where they felt the most support in their drawings of the data science community. Even though I did not explicitly ask what characterized supportive relationships, most people still provided a reason as to why they felt more support from one group or individual vs another. Participants explained that supportive relationships are ones in which: 1) they feel as though their research is understood; 2) they feel solidarity because of shared experiences or culture; and 3) learning or teaching occurs.

Interest and understanding in research

The first aspect was also the most mentioned. People felt supported by when they felt others understood their research or the struggles and goals of their research process. Researchers from both faculties shared this idea. One PhD explained that when they have a problem in their research, they feel the most support from those who can address their concern. For them, this came from a lab group of people working in the same domain/discipline as them. A professor similarly felt supported by a community of people working on similar research: "most likely I get the most support from them, because they actually understand what I'm talking about, what I want to achieve." They felt supported by people who understood their research's meaning. This feeling was shared by another participant who explained that their supervisor "was actually extremely supportive because they really think more about [my research] topics." This same person also explained that they felt unsupported by other supervisors, "partially because they were not interested in these topics, partially because they were considering it's not computer science."

Feeling as though your research is understood was important to people in the community who did not feel supported in certain relationships as well. Another participant also pointed out that a lack of interest from others made them feel less supported by the community. They explained that at least from other professors within the university, "I honestly don't think people are really interested in what I teach."

Shared experiences

Solidarity through shared experiences was another aspect of supportive relationships. Two participants expressed that they feel supported in life more generally by people with whom they shared a cultural understanding. Both were international and non-white, and observed that they tend to form groups with people from the same part of the world or with a similar cultural background. One explained that "in Delft, we kind of have our family here because we have the same culture, so we can share easily." Although these students did not talk about these relationships in terms of the data science community or research, identity can still influence these indirectly though our personal lives.

A few students described "their people" as peers in the same master's or bachelor's program as them. Some of them talked about meeting good friends in their classes or while doing projects. These shared experiences provided an opportunity to meet and also to grow closer.

Young professionals also talked about feeling supported by other people in a similar position as them. A master's student felt the most supported by others in the same "bubble" of academia, because "they did understand me, they did understand the struggle that I had in explaining things and also explaining these things to myself." For this student, they felt supported by others who also struggled with communicating academic research to family or friends who were not in academia. A few PhDs also talked about how "PhDs still feel a little in the same boat." One explained, "PhD students are usually much more supportive than the professors... they are asking similar questions and even those that don't work on these [same research] topics... in their personal life, they actually think a lot about these questions." Another participant wondered, "I don't know if this is a question of age or position, role, but usually people closer in age to me were more open and more supportive." This was again connected to feeling understood. These PhDs felt support from other PhDs, because even if they were working on different research topics, they had similar processes or were open-minded to learning about the research of their peers.

Teaching and learning

Learning and teaching were the third aspect that participants described in supportive relationships. This characteristic of support was described by PhDs, postdocs and professors from both faculties, but not by students. Participants described this aspect as present in different types of relationships. Two participants felt their supervisors were supportive because they encouraged them to explore different opportunities and were "asking me all these questions and teaching me a lot for sure." One participant explained they learned a lot from friends and family because they "not only support my research, but also taught me a lot about future careers." A professor also explained that they try to support students by paying attention to their interests and acknowledging their talent, actively learning from them. "I work with very talented students. Of course I cannot know everything, but I keep my ears quite open to see what students are interested in and see how I can support them because I'm not there like doing the research. It's them. So if I can be a facilitator or enable talented people to be successful and fully explore these questions, then that's the strategy that I follow." While most of the participants felt supported by people who taught them something, this professor also provided support by *learning* from students as well, not just teaching them.

4.5.2 Barriers in Relationships

There were a few instances where participants explicitly expressed a lack of support or difficulty in forming relationships. Often, these feelings confirmed the characteristics of supportive relationships mentioned by others above. For instance, one student of color explained that they struggle with asking for help from supervisors or other students, because they have had experiences where "I talk to someone and they're not listening or they're not trying to understand what I'm saying, and they're just trying to correct me." This participant felt a lack of support (and was discouraged from seeking support in other relationships), because the people they've worked with did not try to learn from or understand them.

Two other participants pointed to disconnects between certain groups in the community by describing a lack of support. One person felt unsupported by professors in EEMCS because they chose to work on a project outside of the university. "The professors there, they told me, you should have never worked on it. Like, it's too political. So I mean partially it's the individual professor thinking that, but I can imagine that it might bee the kind of attitude the faculty is promoting." This participant felt as though the culture within this faculty was in conflict with their own values. Another partcipant explained that "the data regulation guys, they aren't that clear about what are researchers doing... They're responsible for the society, for normal people. They are not responsible for researchers. I think that's the point, because they don't care about us." They highlighted a disconnect between researchers and others in the community working more on data management and regulation. This was related again to the feeling that their research was not understood by this group.

While these participants felt a lack of support, others pointed out structures or disconnects in the community that make it difficult to relate to or connect with other groups to begin with.

Hierarchies and structures

One master's student explained that they found it difficult to connect to professors at TU Delft, "to be honest, I'm a bit afraid of them." They reflected that this feeling was partially because they came from an educational system that was much more hierarchical with greater distance between roles. However, it does provide a potential explanation for why many young professionals feel supported by people in their own peer groups.

The hierarchical structure in academia was only mentioned by three master's students. They shared ways that the hierarchy impacts their relationships. For one, this manifested as feeling greater distance between themselves and professors. For another student, this structure made them feel intimidated by their supervisory committee. This feeling occurred despite that though their committee would "say 'we' are going to do something instead of 'you should'" to actively try and reduce this barrier. The third student referred to the impact of hierarchies more generally. They explained that we need to be mindful that hierarchies exist to use them in a correct way: "just recognize it and then you can use that power, then you use it to kind of elevate the other people or to put yourself below them."

Hierarchies made it difficult for a few students to connect with professors or those above them. For one professor, the structure of the university also made it challenging for them to connect with others. They explained that they don't feel too connected to the university because they don't feel that their research is understood. "People, you don't even know what they're working on. So organizing some kind of events where you can learn what people are doing." Many people mentioned that they were able to meet collaborators and supervisors at conferences or events.

One participant explained that they met their current supervisor from a different faculty at an event; they started working with this supervisor because they shared more values and research interests with them than people with people in their own faculty.

Diversity and difficulty connecting

In a few of my conversations about identity, people expressed that being different from those around them sometimes made it difficult to relate to others. While some participants described feeling supported by people with whom they had shared experiences, these participants felt it was difficult to connect to others because their experiences are not shared.

One participant explained that being a woman in computer science and an international at TU Delft made it hard to relate to their peers: "sometimes, they try to look like they understand my problems, but they actually cannot because they have never been in this situation." Another participant who identified as a POC described the feeling of being different as lonely: "there's always people surrounding you. So you're not alone, but you're lonely in that sense. You're in your experience, and that's because you're the only one that's different, and you're the only one that understands you."

These two experiences, along with how participants described shared experiences as part of supportive relationships, seem to suggest that diversity can make it challenging to connect or feel supported by others. As one participant from TPM explained, "I also don't feel very well connected because I'm at a faculty that is very diverse. So most of the people do not do related things." Another participant similarly pointed to this challenge, explaining that when the faculty is "very diverse in terms of topics... it can be difficult if you are trying to find your place."

However, many participants also expressed both the importance and benefits of having diverse relationships. One person explained that "working with people from different backgrounds is helpful for me" because it challenged them to reflect on their own biases. Another participant involved in a cross-faculty lab group explained, "I enjoy the communication in [this group] because we are doing different research which are not that related, but it's still interesting." Others recognized the benefit of having more diversity across the entire community. One participant, in reflecting on who to consider as a data scientist commented, "I do think also that diversity and inclusion is important in this. So if there is someone also that is not necessarily, is not the label of data officer, but is interested in such kind of community, generally, those people brings a lot of energy, a lot of maybe also different perspectives."

4.5.3 Relationships: Discussion

Participants mostly described how they felt supported in their research endeavors. This makes sense, since I asked about relationships within the data science community at TU Delft – which is a research community. Yet, very few people talked about feeling supported as a person by this community. I only made a distinction between being supported as a researcher and being supported as a person in one conversation with a professor. When I asked if there was support for them socially, they answered, "No, there's not."

This made me think back to the personal values that participants shared (Section 4.3.1). Although only a handful of people talked about valuing respect and open-mindedness, these values had a definition similar to how other participants described supportive relationships. Being respected "for the work that you've done and are doing" means that people understand your

research and its importance. Being open-minded "to the feedback and also to the experience of other people" sounds a lot like teaching and learning. To be open-minded also means teaching and learning from others, even if the experiences of others "don't resemble our own."

This last part was less evident in the relationships participants described. There was a tendency for participants to describe support from people of same level, people in close proximity, and with similar experiences. hooks (2003) writes that even when we value diversity, we have an inclination towards sameness. A similar pattern showed up in reflection opportunities; most informal reflection happens between people in similar positions. It is unclear whether this is the result of or the cause for feeling more support from people in similar positions, but there is a connection.

A few participants also shared with me how their experiences as minorities made it difficult to relate to others. While this doesn't necessarily mean that the community is not open to their experiences, these participants experienced a perception that others could not understand where they were coming from. This perception emerges through the interplay between representation and recognition. For instance, the student who found it difficult to ask for help felt that way because they experienced not receiving help before. The way someone treats us can influenced our own internal recognition of how we fit into a group, which in turn can influence how others see and represent us. This interplay is likely the reason why a few participants at TPM felt it is difficult to find their place in a (disciplinarily) diverse faculty.

Supportive relationships can also be a way to reconcile this conflict between our internal recognition and external representation. Through teaching and learning and exchanging information about our different perspectives, we can experience more understanding between different cultures, identities, positions, or research areas. Teaching and learning is also a way to use privilege within hierarchies. The professor who described keeping an ear open to what students are interested in is actively creating environments in which both teaching and learning can occur. This is a demonstration of the reciprocity that Kirkness and Barnhardt (2001) emphasize. Hierarchies are felt most prominently as a barrier to forming relationships by those at the bottom of the hierarchy – students. Students also feel less capable of initiating opportunities to reflect with others. The invitation from a professor to learn from their student is a way to empower students to feel more capable of forming their own supportive relationships.

4.6 Resources and Processes

When I started the interview process, I was curious about how people access resources in the community, and who has influence over what transformations and processes are going on. As someone from outside of data science, I was also constantly learning throughout this process. My initial list of resources I presented to people was short, but it grew with each interview as I learned about the changes and processes available to researchers. Towards the end, discussions around processes, transformations, and resources became much more interesting as I also had more information to provide.

The initial list I provided is reproduced in Table 4.4. It was intentionally general since I was discussing with people from multiple faculties and wanted to prompt conversation rather than to create an exhaustive database. From this starting point, I could reflect on what people choose to add to the lists or to comment on. Especially in later interviews, I would also probe the participant, asking if they were familiar with a particular process or resource. Because of the unstructured way I approached these conversations, individual responses cannot be directly compared. Instead, I consider how additions or revisions to my list reflect certain values or desires for the community at TU Delft as a whole.

Table 4.4: Initial list of resources and processes/transformations

Resources	Processes
• Funding • Mentorship/guidance • Course materials • Work and study spaces • Number of students	• Research projects • Curriculum developments/changes • Labs or research groups • Personal connections and interactions • Development of policies or practices

Events and networking

A professional who is relatively new to the TU Delft explained, "you know, there's a lot of events that happen here." Both professionals and young professionals brought up events that they had either attended or with which they were directly involved. Some of the events that came up were specific to a domain or area of research, such as an event about natural language processing or lunches hosted by MONDAI, the cross-faculty AI hub that recently formed at TU Delft. Other events mentioned were hosted outside of TU Delft but still relevant to the researchers I talked to, such as conferences or public Meetup events on data science. There were also a few events described that were strictly internal to TU Delft, such as the 100 Days of Data Education. Two PhD students described events hosted by their faculty or their lab group designed specifically for new PhDs to get to know one another.

These events were described as places to network. One professor explained that they liked the type of events that invited industrial people from specific domains or topics, "because they have presentations and you could learn what people are working on." A PhD also described the importance of adding Meetup events since, "it is a process or transformation because it does connect people. It might be a resource also, because it connects people. But it carries around the conversation on these kinds of things."

These types of events also received comments in later interviews such as, "I don't know what

that is." The two instances where professionals expressed this, it was unclear whether they were curious to participate in the event or only wanted to know more about it. Regardless, there are events that may have relevance to their work of which they are unaware. A third professional even explained that "in the processes and transformations, there's a few more things that I probably have now discovered just looking at it and learning from you, so that's been fantastic."

These comments came entirely from professionals. This may be a skewed observation, since I talked with more students and PhDs at the beginning of my process and more professionals towards the end. Regardless, interest in these events seems to be connected to levels of expertise. None of the learners I talked to described events they had attended or knew about, and they also did not comment on events already listed. Although one learner pointed out that "leisure time within the campus" is something they appreciate having available.

Mentorship, support roles, and people as a resource

Professors and students both commented on mentorship and support roles, mentioning that they are important. One professor suggested that I add "counselling and buddy programs" that are meant to help students and PhDs. A PhD also suggested adding trust-people to the list, "a person that you go to if you have something personal that you maybe wanna somehow share to the whole group or the head of the group... then you have this confidential person." This was a role present in their lab group that they had access to directly. In thinking about what types of support roles are important for the community, participants thought of both traditional roles (like supervisors) and roles aimed at supporting people, rather than just research.

Data stewards were another role that was mentioned as a resource, but as with deciding whether to include this role in the community, not everyone agreed on whether this role was a valuable resource. One professor felt that "having meaningful support from data stewards and scientific programmers would be very valuable." However, a different professor commented on this addition, "I've heard of it, and I know the students have to do this. But I don't know if it's really useful for them."

Besides reflecting on specific roles, there were also two people who directly commented on "number of students" as a resource. One professor explained that it is more about the number of capable students, because "some students need way more guidance. So then it's less useful for you in the sense that you don't really get in this interaction." This comment ties back to supportive relationships being those relationships that you can also learn from. A student also commented that "I think it's not only about numbers, it's composition that is important." This student felt that having a diverse student population was a valuable resource for them, because it introduced them to different perspectives.

Changes to university policies and practices

I included the development of policies or practices on my list because I suspected that these larger, institutional changes were happening. As an outsider and also a student, I did not come into these conversations knowing about any specific changes. This subject was a particularly interesting one to talk about with participants of all levels.

I very quickly learned about a major development happening in EEMCS, that they are creating a new data science master's program. One professor explained, "I think everyone is now participating in this." I asked if students were involved, and they explained there is a "selected group

of students, also diverse, that help in the design." A professor in TPM also commented on how involved students were in shaping faculty-wide policies: "there is one representative of the student in this [education-related] meeting actually, so they do bring their opinion to the table, but it's kind of a representative. It's not necessary, might not necessarily be the widespread opinion."

Although some students are involved in these faculty-wide processes, the students I talked with had differing views on how much their input was valued. Two master's students talked specifically about the curriculum feedback forms provided at the end of courses, which are a widespread way the university receives input from students. One student in EEMCS felt that "from our feedback, they at least look at it..there's some curriculum changes." They felt these changes in curriculum, including the development of a new data science master's program, were valuable and a way they noticed their faculty adapting to student opinions. On the other hand, a student in TPM told me, "Oh I never fill those [feedback forms] out. I don't think they're really useful actually." This student had the opposite feeling, that their input was not taken into consideration in curriculum changes.

A few participants also felt that they had little influence over policies that affected the university at a wider scale. One professor explained, "I don't think I influence the policies a lot, at least the global ones not much." A master's student also commented that, "I think processes and transformations is, general public is usually not involved." They felt the university should have greater transparency in decisions. On a smaller scale, one PhD felt that they were able to influence the development of policies or practices, at least within their lab group: "I feel that we can talk about stuff like that."

Lab groups, Teams groups and projects

Another process I initially included in my list was research projects. Many participants also talked about the groups they were a part of as resources (or as both a resource and process). The groups they described varied in both purpose and also accessibility.

One professor mentioned that they were part of a Microsoft Teams group about machine learning. They were added to this group by a colleague who thought they might be interested. This type of group was both not discoverable (I was unable to find anything about it publicly available), and not well known. A different professor from the same faculty commented, "I don't know what that is. . . I'm not aware of that." However, the idea of these types of groups is that they are open to anyone interested in the subject.

PhDs and professors also mentioned lab groups they belong to or lead. Some of these groups are directed by and contain students from one faculty, while others were established cross-faculty (either independently or as part of the TU Delft AI Initiative, which established cross-faculty labs groups around different AI themes). These types of groups were discoverable if you know about them – I was able to find websites of the various lab groups participants mentioned. However, they were not necessarily open for anyone to join. One professor explained, "I was a bit sad that they didn't have more proposals" to set up AI labs, as they would have liked to be part of one. When I asked if they knew whether they could join an existing AI lab, they responded, "no,no. As far as I know...my understanding is that you cannot join. But I'm also not 100% certain."

As with networking events, students did not talk about lab groups except when I prompted

them about it specifically. I asked one student who was part of a research group whether they felt they could use that group as a resource. They explained, "I like the idea of research groups and I like the idea of [my group]. But then, I don't know that if I were to ask for help that I would get it. So in that sense, how valuable is it?" Despite being part of a research group, this student did not feel like they could access the value of it, at least not fully. I also asked a learner who was not yet part of any research group whether they knew how to get involved with one. They answered, "not really, because I know [one professor] sent us an email, but other than [that professor], maybe not all professors sent us, give us the access to join." They went on to explain that they would like if they received some kind of newsletter, "you can put some informations in an email or just post it in a group...So make it more accessible for us to reach out." This student pointed out that having more information would make these types of resources more accessible to them.

4.6.1 Accessing Resources and Processes

The ways that people described accessing resources or processes typically involved 1) learning about it through people or connections or 2) actively searching for it yourself.

A lot of participants described learning about events or joining groups as a result of direct connections. A student talked about how their classmates often share "course materials, you maybe ask somebody, ohh do you have access to this?" Another student talked about getting support for their start-up by talking with "my classmates, or the classmates from the year below." These students talked about accessing different resources through their close peer groups, a characteristic that I also included in a relational community representation (Figure 4.6).

There were also two participants who talked about "a snowball effect" in which they contacted one professor and were put in touch with another professor and so on. For one student, this helped them find a research project to get involved with. For the other professional, they described this as a way they made more connections within their faculty. A professor emphasized the importance of individual connections for getting involved with multiple disciplinary communities or groups: "bridges or connections from what I've seen in my experience, they are made by someone that has a personal interest. They say OK, I'm in this community, but given your profile, I think you would be fit also for this community." In these examples, individuals took a proactive step to connect participants to another group or individual, which in turn lead to something valuable for them.

However, three students talked about needing initiative to access resources yourself. "If you are actively enough to find a mentor, you will get one, but the mentor will not come to you automatically." Another student explained that researchers "should be the one kind of reaching out to the community itself. The human capital that you have around you is the tool." When I asked a learner whether they knew how to get involved with research projects, they explained, "I believe that if you are searching then you can you can definitely find something."

A common sentiment people shared about resources was, "I have enough understanding of where to go for these." However, there were some resources that were not easy to access because they were limited (rather than because of a lack of knowledge). One professor explained that "funding is always scarce, you have to apply for it." A student felt that study spaces were limited: "the library is full every day, even if it's week one or week 10." A PhD explained that "we need GPUs very hungrily, but we don't have a lot in TU Delft or in our department." This was echoed by a professor who felt that "you always have to battle against the hype

technology... You can either join them, but then you have to compete with a huge community and rely on large computational resources which we don't have." People pointed out limitations to resources that they desired. PhDs and professionals did not mention limited study spaces, for example; likewise, students did not talk about the scarcity of funding.

However, people from all positions expressed concern about data accessibility. One learner told me about an online resource that provides training for data scientists, along with accessible data to learn with. They felt it was important I include data accessibility on the list. A young professional highlighted that accessing data in their domain is dependant on having certain knowledge: "it's very difficult for people to use this data unless they have very highly specialized knowledge." Other young professionals pointed out that the type of data they use in their research can either make more challenges, or make it an easy resource to access. One of them explained, "many of the data we need to buy...So it's somehow like resources are limited because I cannot use their data so I cannot validate my measurement criteria, so I don't know whether this methodology is acceptable." However, the other participant explained, "using open data sources makes it very accessible to me." Finally, a professor highlighted why accessible data is valuable for the community by suggesting that we need "incentives dedicated to making data accessible and well documented."

Access through the internet

The internet and ChatGPT were both mentioned as ways for people to access data or join in data science processes. They were discussed almost as equalizing forces because they made data and data science more accessible to society. The PhD who described attending meetup events explained, "I think you have this website which is called meetup.com. This data science meetup is on that website. So basically anyone can join." A student described a platform called Kaggle that provides competitions so "many people around the world that are curious about data science or AI" can learn about these processes through an open competition. Another learner felt they can get data they might want through the internet, and as a resource, it is close and easy to access. One PhD explained, "when ChatGPT comes, the world is change. Industry is even faster than academia. It's a huge change for all the academic researchers, so more social thinking [about] how the research can contribute through the industry to the whole system." They felt that ChatGPT was a process causing researchers to reflect more on how they can contribute meaningfully to society.

While students shared generally positive examples of using the internet, two professors had a much more critical take on these open resources. One gave the example of sklearn, an internet package where you can download algorithms from the internet. This professor explained that "I am very surprised that there are no antidiscrimination routines there." The second professor spoke from personal experience about using the internet when you are unsure how to program something. "You don't just take the code from the internet, and I discourage people from doing that because I also burn myself on this." Only one young professional described doing this, explaining that when they have questions about the limitations of their methods, "mostly I search on the internet, asking Google." This same participant had a desire for "more communication between mathematical guys and engineering guys." Perhaps because they didn't have many connections to researchers with knowledge about data science methods, they instead turned to Google with their questions.

4.6.2 Resources and Processes: Discussion

Conversations about resources and processes went in many different directions. I approached this part of the discussion with very little structure. The questions I probed participants with were inconsistent, changing based on who I was talking to and what was on the resource list at the time. I would have liked to be more thorough with updating my resource list; after interviews, I added the things that I recalled from my conversations, rather than listening back to the audio file to catch all mentioned resources. There were also two interviews in which I did not have time to ask about resources at all. For these reasons, I cannot say anything conclusive about whether certain groups do or do not have access to certain resources or processes. Nevertheless, these conversations prompted interesting reflections on what makes people feel like they do or do not have access and what resources or processes they might want more of (I will discuss ideals and goals for the community more thoroughly in Section 4.7).

The culture at TU Delft is one in which you are expected to look for the opportunities, resources, or support you might need. This is something I have also experienced during my studies here. While this does not inherently prevent access, different cultures might interact with such a system differently. For example, Western cultures like the Netherlands tend to be more "ask" cultures; the social expectation is that you will ask for what you want and others do not need to offer something to you unsolicited. However, a lot of East Asian cultures tend to be more "guess" cultures, where the social expectation is that people around you will guess what you might need from context and offer it to you. The model for accessing resources is more aligned with ask culture. This may be unintentionally limiting how people from other cultures are able to participate.

There are a lot of things happening at TU Delft and within the data science community. My impression from talking with participants was that there is too much to keep track of, especially when it comes to events. This can be a challenge as well; how do you get word out about a resource or process to the people who need to know about it? For example, I learned about a resource called the Digital Competence Center towards the end of my project.[1] This center provides the kind of data management and software support that one participant felt would be valuable for researchers. Yet, in talking with someone who works for this center, I also learned that they are currently small and unable to take on all of the requests they receive. Although they want to grow and become more visible, there is also a contradiction in that they might not be able to accommodate increased visibility. So while this resource (and I suspect others as well) is limited, it doesn't feel limited if we don't know about it to begin with.

Having knowledge about resources and processes in the community is an important part of access. When people suggested things to add to the list, they typically only mentioned things that they were already involved with or had access to. The exception to this was a few professors who discussed resources primarily available to students, such as buddy groups. Information seems to travel from professionals down to learners, and less often the other way. Participants also perceived this trend, choosing to represent the relationship between professors and students as uni-directional (Section 4.2.2). In Section 4.5.2, I wrote about how students feel the academic hierarchy impacts their relationships and in Section 4.4.3 one student felt like they were not capable of creating opportunities for reflection. These observations make me wonder whether students actually lack information about resources and processes that might be relevant to professors, or whether they only feel like they do not know these things.

[1] I learned about the TU Delft Digital Competence Center (DCC) at an Open Science Networking event hosted in September 2023.

The observation that master's and bachelor's students had less to say about events also indicates differences in access between positions. These types of networking events, despite being open to the whole campus, tend to target PhD level positions and above. While there is a argument to be made that we don't need to focus on in involving students because they are at the university for less time, I also feel there is a missed opportunity to learn from this group of people. Especially considering how diverse master's students are — and the fact that many have worked in industry prior to attending TU Delft — the community loses out on potential insights from these perspectives.

Although these networking events were less targeted towards students, on the other hand, existing lab groups felt less open to professors. It stood out to me that one early-career professor did not know whether they might be able to join an existing lab. And while many lab groups will post on their website if they have a vacancy for a PhD position or a master's student project, I have not seen a similar call for collaborating with other professors. Perhaps this is due to an expectation that professors should aim to set up their own group, rather than joining an existing one. This expectation is also a form of competition in academia (see Section 4.7.4).

I did not know many of the resources and processes mentioned by participants prior to this research. For example, when I started this project and searched "TU Delft data science," I found one website that linked to a few professors working on different themes. MONDAI did not come up, neither did the TU Delft AI initiative, nor 100 Days of Data Science Education, all of which sounded like big initiatives from my conversations. If you don't know what to search for, information is limited.

This also relates to how participants discussed the value of transparency. By their definition, there is not a lot of transparency around accessing resources or processes. It is usually not explicit which audiences events are made for, or whether certain resources are available to all people at TU Delft, only students, or only employeed researchers. A report from 2019 surveying employees at Dutch Universities found that the deliberate withholding of information was the second most common reason why people felt unsafe in their work environment.[2] "Dominant groups often maintain their power by keeping information from subordinate groups" (hooks, 2003, p. 74). While resources might not be intentionally withheld, it is important for us to reflect on what our responsibility is to share information about resources, especially when they are so often accessed through others. This consideration is especially relevant for groups organized around themes or topics of interest within data science. Who is invited into these groups? Are they searchable by those looking for them?

Participants also expressed a desire for more ways find information easily, such as through emails or newsletters. A few participants also emphasized that physical spaces matter, both for events to network with other people, but also because you can see various projects or information on display. It is interesting to also consider the ways that the internet and ChatGPT are shifting the landscape of data science. Participants pointed to how these resources are expanding accessibility and bringing data science closer to the people. ChatGPT and the internet have the potential to be a conduit for redistribution of resources and allow for greater participation. Young professionals and learners felt this most – yet professors felt more hesitation to consider these online resources as reliable. This hesitation reinforces the idea that data science experts understand the implications, especially negative implications, of their work more than learners or young professionals (Section 4.2.1). However, this hesitation also reminded me of a comment

[2]Report, published in 2019 by the FNV and VAWO, surveyed 1,110 employees at Dutch Universities, including support and academic staff. [Accessed: 16 Oct 2023]

from a young professional. They felt academia missed a major opportunity to make their educational resources widely available to the public after COVID-19 forced education to shift online. They felt that part of the decision to not make courses available digitally was due hesitation from professors, "because then professors feel threatened by the fact that people are not attending the classes...they're not attending because they don't see it as something valuable more than having a sheet of paper in which everything you say is written down."

While I cannot speak for the professors I talked to, perhaps there is a similar resistance in the data science community preventing wider adoption of open publishing or acceptance of tools like ChatGPT. But rather than responding to this change in the community by keeping resources or expertise to ourselves, this could be an invitation to reimagine our roles as data scientists. This might create the space for data scientists to see their "expertise" as not just manipulating data, but actually understanding on a deep level what you are trying to accomplish with the data and how to communicate that. Perhaps this also allows the community to more clearly acknowledge the ways learners and young professionals can be experts in their own right. Sometimes an expert is the person who know the right questions to ask.

4.7 Community Movement

When I was initially defining my research questions, I kept returning to the idea of a center — the point at which a community rotates or gathers around. My research design is concerned with how to center perspectives that are underrepresented in the community, framing this intervention as a way to meaningfully include a greater diversity of people in data science. This is based on many assumptions that I challenge and reconsider in this section, starting with whether the center of the community is actually where important things are happening.

I learned a lot about what is happening in the data science community at TU Delft by asking about the processes and relationships I've described in previous sections. I also used the drawings as a starting point to ask about the center of the community and where there is momentum for change. People also shared throughout the conversation about limitations or disconnects, structures and regulations, and their ideals for the community. I consider these all to be forms of community movement, both preventing and enabling things to happen.

4.7.1 Community Center

In seven of my interviews, I explicitly asked participants about where they would put the center of the community. I gave no additional explanation of what it means to be the center of a community and almost all of the responses were unique answers. The lack of consensus was partially because people were thinking about different types of community organization; some were thinking relationally, some institutionally, and some in terms of disciplinary or domain boundaries.

Two students answered this question by placing people in the center. One student explained they would consider professors the center of the community because "they both teach the TU Delft students, but also do their own research with other people... so they have connections with almost everybody." The second student put themselves in the center, explaining "I don't want this to sound arrogant, but I am like the center of my life." They felt that the community is "more bottom up than top down," and from their perspective, students are a "main character" in the community.

Two responses named a center based on an institutional perspective. A PhD simply placed academia as the center. A master's student, while they acknowledged that "academia has an impact on like the world," instead put the center of the community around what they called "home." They explained, "that is the community that we are serving."

A similar response came from another PhD considering the center for their research domain. They felt the center was "a whole vein of people." But rather than the people being what held this group together, a shared research purpose is what they gathered around: "the topic in the middle, it's well-being of people." Two researchers, instead of considering the center as located within one domain, placed it at the intersection between three domains. One PhD felt the center of the data science community should be a combination of engineering research; mathematical, physics, and theoretical computer science; and regulation and policy. They affiliated the last two groups with the EEMCS and TPM faculties respectively. A professor similarly positioned the center, but in more abstract terms. They described the center as "an overlap between the domain, the technical part, and also the philosophical part. So it's like domain, philosophy and methods."

4.7.2 Momentum for Change

Irrespective of where the "center" of the community was placed, momentum for change came from many sources. The student who placed themselves in the center of the data science community talked about movement occurring from the bottom up. Other participants also mentioned movement in the community when they discussed initiatives or changes happening.

Aligned with the idea of bottom-up organization, participants from both faculties mentioned specific people who are pushing a lot of projects or initiatives. Professor J. Yang was mentioned for "an effort to create some kind of community for people working on NLP." Professor G.J. Houben was noted for being the "Vice Rector Magnificus of AI at the university" and leading campus-wide AI initiatives. Professor C. Liem was discussed as pushing for conversations about "more critical sides of AI", but that "she seems to be doing a lot individually." There were also a few university-supported initiatives mentioned that host a lot of events, such as the House of AI (MONDAI). One professor explained that "there is no one single main community and I have also sometimes the impression that those communities are like kind of grassroots initiatives. So they pop up because someone has an interest and someone jumps on board." I asked this professor if certain micro communities have more influence over the overall community. They responded that especially in research, "it's where the money is that people are pulled."

Specific people, initiatives, and micro-communities were only discussed by professionals as sources of momentum. On the other hand, two young professionals felt that momentum comes from the younger research community. "Younger people tend to be a little bit more, like, yeah, I want a change. I wanna see things different." The second young professional emphasized that "you need the energy of people who are learning" to create potential for change. They described academia as a "politicized environment for learners and young researchers" where they gain exposure to the professional world, but also are in a stage in their career where their decisions "will not have consequences on their lives necessarily." This means they "can still express very strong opinions on it and they can have strong ideals, they can reject the system."

Academia was also described as an environment that generates momentum by two TPM students. The first explained, "TU Delft is a platform for us to interact between the expert and me, like the unknowledgeable pupil. So I think the interaction is built up in a good way. They are in a nurturing environment." The second student felt that universities "set the tone on how data science, you know, how we should go about it" and also "train those who will be part of the community later on." For these students, academia plays a role in shaping the practices and standards of the community through education.

Some people felt that the momentum for change was coming from outside of academia altogether. Three people in EEMCS felt industry was a major driver for change in the data science community. A PhD commented that the trend towards using big data in industry pushed many of their friends from other disciplines into data science. Another PhD felt momentum was partially from academia and partially from businesses and hobbyists with a genuine interest in data science. They described hobbyists as the people who were "most like entrepreneurial or the most, in that sense, forward thinking." Similar to how ChatGPT was mentioned as making industry faster than academia, this PhD hinted that hobbyists also operate and innovate faster than academia. A professor explained that this industry innovation influenced the faculty's decision to create a separate data science master's program: "it's all driven by the improvements that big tech companies make." They felt that momentum for change came from comparing to "other universities that are creating programs that are called something artificial intelligence, or

something with machine learning." These researchers all felt industry was pushing innovation in data science, influencing researchers and academia to adjust their thinking and education to keep up.

On the other hand, people in TPM talked more about this push for change as coming from policy and government. One PhD talked about how the introduction of data regulations has shifted how people think about their research: "we need to consider [data privacy] and I think algorithms need to be evolved into a new stage." Another researcher explained that in TPM, there seems to be a lot of curiosity and momentum about "how to have more impact, whether it's about activism and NGOs or policy making, etc."

4.7.3 Positive Changes and Ideals

I asked some participants about what they would like to see changed about the community. Many also talked about changes they already saw happening or that they think the community would benefit from without prompting.

Changes in education and curriculum: There are already some big educational changes happening in the data science community with the creation of a new master's program. But a few participants pointed to smaller changes that they felt were valuable as well. For instance, the creation of a course on ethics in AI (offered in EEMCS) or expanding on the foundational programming skills taught in the first year of a TPM master's. One student explained that there are "some transformations going on in terms of like quality of teaching, which I think are really important." Another student pointed out that the curriculum, "here it changes every year or every two years, and it's very good."

More cross-faculty events: One desire for the community that was shared by professors was more events to bring people together from different faculties. They described these as "some spaces or some informal events to bring people together" where "they have presentations and you could learn what people are working on." One professor explained that "all the connections that I made was via some events. All this. That's why I think it's important to have more like cross faculty events like or even theme-driven [events]."

Besides events, cross-faculty initiatives or communications were also desired. One person explained that forming AI labs "was a good effort... because these labs were created between people from across faculties." A PhD felt that the community would also benefit from "more communication between mathematical guys and engineering guys."

More communication and collaboration outside of academia: Beyond the academic community, a few students also talked about a need for more communications and events that involved the public. One student felt there should be more general public events in which professors or researchers "have to make [their research] accessible not only to the academic community, but to the people." Another student felt we should "see like how expert and the society other than data science community can interact with each other." They felt this communication was important so that "we don't have to make an unnecessary debate in the near future about that data, and how can we see data." This comment was referring to debates on social media that can arise from misinformation.

Two PhDs also desired more collaborations with institutions beyond academia. One felt that there were many "industrial opportunities for like data science, they can make more connections" to academic research. The other PhD felt that more connections between research and

government would allow them "to be more involved in more practical things of my research."

Shifting values and regulations: Some participants pointed out positive ways that the community is shifting its values. One young professional explained that "people are taking more consideration, taking more care" towards data privacy compared to when they studied here previously. A professor explained that journals are pushing for more data accessibility when publishing and they "suspect it might be the same for TU Delft." A PhD confirmed this shift to more open and accessible publishing, noticing that the boom of machine learning pushed "other fields to also make it open, their stuff." Open science and open research are also becoming a larger part of TU Delft-wide initiatives and strategy.

Despite these shifts towards more data privacy, greater data accessibility, and more openness, participants still emphasized a need for regulations to solidify these values. One person felt that industry especially still needs more "strict guidelines [about data protection], very strict laws that should be applicable to not only big companies, but also medium companies and small starting companies." A professor and a PhD commented on the need to improve standards and incentives for sharing data in an accessible and understandable way. The PhD felt that researchers should take initiative as individuals to share their data, but that "the university should enforce our researchers to propose open data and use open data" as well. The professor felt that there could be more training and clarity about how to meet existing standards about publishing accessible and reusable data.[3] Another professional also felt that developers working on AI need to reflect about problems of fairness and bias more.

Diversity and representation: A few people from TPM talked about the need for greater diversity, inclusion, and representation in the community and university as a whole. One participant pointed out that already, "there is a lot of communication, lot of openness to speak out about, attention put into topics such as diversity and inclusion." Others felt that it was important to "have a serious discussion about the composition of your students" at the university. As a few people pointed out, "I know the student population, which then becomes the PhD population, which eventually becomes the professor population, is highly unbalanced."

Although multiple people pointed out the importance of having a diverse community, very few were able to give concrete ways to make it more diverse and inclusive. As one professor explained, "we try to get a diverse student population, but it is not easy." A student commented that we need to "make it more comfortable for like other people of color to join in." A PhD proposed the idea of deliberative democracy to increase representation in positions of power, because "if we give the power kind of randomly based on your demographics, it also helps fighting against this imbalance because maybe you won't go for these positions with the power" if you do not already feel represented in those positions. These two actions both centered around the idea of getting people who are not already represented more involved and comfortable in the community.

4.7.4 Limitations and Disconnects

I was also curious to learn about the limitations or disconnects people observed in the community. Some of these limitations were unsurprising; they are things I have also witnessed or experienced at TU Delft. However, there were other less obvious barriers to community movement and interaction that participants pointed out.

[3]The professor referred specifically to FAIR Guiding Principles, which stands for the Findability, Accessibility, Interoperability, and Reuse of digital assets.

Time

The most common limitation mentioned by professors was time. Time affected everything from collaborations to quality of work to teaching. One professor explained that they want to work together, but "everybody is overloaded with work, so it's like we don't really have time to work together. It's a bit sad." This professor, along with another professor felt that time also constrained their ability to teach certain data science concepts or to make their lectures more interactive. One explained, "I get like 2-3 lectures and that's my time." Time was also mentioned as impacting quality of research. One professor who values rigorousness in their methods pointed out a problem "that sometimes you need to quickly do something, quickly implement something, and quickly get some results to see if something is working and you might not have time to check everything."

Students and PhDs also felt the pressures and limitations of time. One PhD explained that they weren't able to collaborate on a certain project because "if I do that connection now, I think that I will be late in my PhD as well. So I am, just for a matter of time, I'm deciding for now, not to do the connection." For a learner, balancing priorities in limited time caused them to "be overwhelmed by all the things that I have to do here – and also work, and also dating life, social life, talking to my parents."

Time pressure not only pressured individual decisions and feelings, it also had an affect on others. One participant explained how the busyness of professors impacted them: "I find that here it's just like almost impossible sometimes to get even the three supervisors into one room." A professor explained how this time pressure can also be detrimental to the community as a whole: "professional level people are very busy in general, and being active – being part of this community – also requires time investment."

Competition

The time pressure people feel in the community is also connected to a culture of competition. One professor explained directly that the time pressure to "do something fast and publish fast" comes from a lot of competition; "you can't get away from it." Two participants also highlighted how the difference in pace between academia and industry can create a feeling like we need to "keep up" with industry. One person called this "a big problem with the data science community, in which it evolves really fast... Like I don't know how academia in its current form can keep up." The other person also pointed out that research "it does take time, while in my eyes entrepreneurial people tend to be more like, just keep going." While neither stated outright that they feel competition towards industry, others shared about how industry influences the direction the field moves in as a whole.

Two people also experienced competition as a lack of fairness. One student felt that it is problematic that they are granted (upon graduation) the same degree and title as other students who have invested less time or work into really learning data science methods. "I have to compare myself with people who have my same degree and my same title, and I know for sure that they didn't put the same time. But a person on the other side might not know it, so for me, it's a problem. like that is a big problem." One professor explained that the trend to optimize for predictive accuracy in research sometimes leads to unfair competition in publication. "Sometimes it's very annoying. Sometimes there are papers written that actually contain bugs in them, such that they get a higher accuracy and then you have to compete against people who cheat. That also happens."

How research is valued was another aspect of competition people pointed out. One professional pointed to a disparity in how funding is awarded, explaining "it is much easier to well write a grant about some AI development than writing a grant I guess criticizing AI or like reflecting about potential limitations." Another professional experienced this in their own research, describing it as "not a very sexy topic." They felt that more recognition for researchers who are working on publications about the implications of assumptions in data science research would help. As a student pointed out, you have to have "many papers published in order for you to rise in the ranks. So the goal should be different."

Adapting to the system

A limitation to change in the community is also the difficulty of individuals to go up against a system as big as academia. For example, multiple participants pointed out that we need different metrics or incentives to encourage more open science and open publication. I asked one participant what prevented the shift to open science, and they explained: "if I would answer very in an easy way, I would just say it's the fault of the journals...because publication is focused on KPIs [key performance indicators]. Researchers will not adapt themselves." Researchers are not incentivized to act differently by the system, and this participant felt that researchers will not change in a way that does not fit the system.

Another participant highlighted the difficulty of trying to go against a system as an individual. Although they saw that a lot of people in academia pushing for change, "the only problem I see is after a certain point, people are so hammered by the system that you just give up. You adapt to it...For individuals, it's really hard for them to make change." A different participant emphasized why this adaptation is frustrating for those who feel like they fit the least within the university: "I feel like responsibility of change lies with someone who is like a victim of the system...Like, how should you change instead of like—-why should I be the one to change though? Like why can't everyone else change? I haven't done anything wrong."

4.7.5 Community Movement: Discussion

Although participants described having a lot of freedom and agency in their research, many of them still noticed that there is pressure to conform within academia. Indirectly, it impacted how people described their own work as not trendy or "sexy" in data science. Sometimes, people felt this made it difficult to publish or get funding because other subjects in data science get more attention. Academic competition is a manifestation of dominator culture, creating a sense of urgency and need to compare ourselves to others. Two individuals felt this as a lack of fairness because they were not acknowledged for their expertise or work. However, this sentiment echos how other participants felt a lack of support when different parts of the community did not understand the value of their research.

This kind of competition also operates through a scarcity mentality, because we feel like recognition and incentives are limited. Being more creative with how we incentivize and acknowledge research, as a few participants suggested, can help overcome this limitation. Franklin et al. (2023) similarly call for a shift away from the trope of "publish or perish" towards more collective forms of academic survival or recognition. Redistribution might be another way to counteract a scarcity mentality by intentionally sharing resources with others. How often do researchers give to others, expecting nothing in return? If we reframe redistribution as a responsibility to give what we have, we can challenge academic competition and create a more cooperative environment, at least on an interpersonal level.

Competition also showed up on an institutional level. This was evident in the justification one professor gave for why EEMCS was creating a new data science master's program. Other universities were creating new programs to keep up with the evolving industry, and TU Delft needed keep up with these universities to "attract students" by also offering an interesting, competitive program. Competition and comparison are deeply ingrained into our academic systems, even through how universities advertise to prospective students.

The goal of attracting more and more students was critiqued by a different participant. They felt that academia's goals should instead focus on serving society. This stance echoed an argument made by hooks about education systems. "Commitment to teaching well is a commitment to service" but "dominator culture pointedly degrades service" by regarding those who serve as unworthy and inferior (hooks, 2003, p. 83). Many of the participants I talked with are committed to doing research that will serve people. They expressed this commitment as a value to contribute meaningfully to society, and as a motivation to tackle real problems they see in society (see Section 4.3.1 and Section 4.4.1). Although most of them did not use the language of "service", they still described centering people in their research choices. For instance, by representing more parts of the world with their data, or through collaborations with industry or government institutions so they could have a more tangible impact on society.

I see this commitment to service as a strength of this community. In their own way, individual researchers are resisting dominator culture and competition by pursuing what they believe is interesting, rather than where there is hype or what is easy to publish. Chin et al. (2022) observed that researchers balance their own interests with what might gain scholarly attention when picking topics. Although opportunity and money were both motivators affecting research choices, only one person was drawn into their research subject because of the hype around it (see Section 4.4.1). So while there is pressure to pursue similar research goals in data science, within the TU Delft data science community, there seems to be a lot of agency for people to pursue research that reflects their own perspectives.

Representation was also evident in how people described momentum and the center of the community. One person felt the center of a community is connected to a shared research purpose. Another person put themselves at the center, while others named different individuals or initiatives who are generating momentum in the community. The variation in responses suggests that there is no one center to the community. Rather, the community is a collection of smaller research and interest groups within the larger umbrella of data science. Each one of these might be considered their own center, organizing around either a shared research interest or a specific person or initiative. The data science community, through this bottom-up organization, is responding and adapting to different perspectives from individuals. There are concrete examples throughout these results of individuals changing community culture by influencing the people around them: a postdoc creating more opportunities for researchers to reflect on their biases, a professor teaching both method and design values in their classroom, a young professional prioritizing representation over data quality in their research, a lab group introducing a "trust person" as they grew to provide members with more social support, etc.

Different people, initiatives, or even institutions have a pull on different parts of the community. Young professionals felt momentum for change comes mostly from other young people, while professionals named other professionals they knew organizing initiatives in the community. Students felt that the academic environment shapes practices in data science, while PhDs with previous experience working for or collaborating with companies felt industry has influence over innovations in the field. These are all sources of momentum in the community, influencing

changes in perhaps different ways. However, people tended to point to sources that were close to their own experiences. This might be an unintended bias; people are likely more familiar with topics or individuals doing work that is relevant to them. However, this bias demonstrates the importance of relevance and being represented for creating a more diverse community; seeing people like us in roles that can shape the community also helps us imagine ourselves in those roles. This is the same justification one participant explained for using deliberative democracy to increase representation. It also poses a question for us to reflect on: how do we acknowledge (and support) the work that people are doing to change the community when it does not directly impact us?

I see a potential solution in how two participants imagined the center of the data science community. They felt the center was (or should be) an overlap of domain, philosophy, and methods. For them, instead of momentum coming from an individual or a shared purpose, data science revolves around different research approaches. This way of imagining the community invites in more perspectives, rather than encouraging or prioritizing one framing to approach data science. Identity impacts our research framing, so to consider the center of the community as an intersection of different framings creates more space for expressing our identities as well.

Going into my interviews, I did not specify what the center of a community might be, nor what momentum in the community refers to. This led to a wide variety of answers. However, if I were to start the research process again, I would be curious to ask participants how they would define "center" or "momentum." I would also want to explore momentum for specific changes or areas of the community, rather than as a general feeling. Specificity could have made it easier to compare answers between participants and assess whether the parts of the community are incorporating many perspectives or only one. From my conversations, it appears that data science as a whole is shaped by and relevant to many different perspectives. However, there are so many micro-communities and bottom-up initiatives, each with their own purpose, that I cannot be sure if the diversity in perspectives is only a result of data science being difficult to define. I wonder, for instance, if I might draw the same conclusions for only the AI community. MONDAI seems to be an influential initiative within AI, as it was mentioned in many of my conversations. There is also a lot more top-down support from the university supporting AI-related research. Does the top-down support lead to more or less hegemony in research goals, priorities, and values within AI?

5 | Design

The underlying motivation for this project was to not only learn about inclusion, but to also facilitate a more meaningful form of inclusion within the data science community. An essential aspect of any decolonizing efforts is action, and I wanted to create something with this project that could be directly used by those in the community. My design is a short, interactive activity for smaller groups (labs, centers, research groups, etc.) to reflect on and strengthen their connections to the larger data science community at TU Delft. The activity outline along with an info sheet is in Appendix F.

I learned a lot about the data science research community at TU Delft from my exploratory interviews. It was difficult to know where to start with my design, because there were so many possible directions to go in. I doubted whether what I chose to focus on was going to lead to something meaningful, useful, or valuable. However, as one participant pointed out, an individual cannot change the system alone (see Section 4.7.4). It is not realistic for me to expect my design to entirely reimagine the community. Instead, I focused on a goal that felt achievable within the scope of my thesis. In this chapter, I will first discuss the the inspiration and results that informed my design process. I then introduce the different parts of the activity, their expected impacts and how the activity might be used in practice. I finish with a reflection on this activity's applicability to other situations, as well as limitations of my design process and the design itself.

5.1 Defining a Design Goal

With my research questions (Section 1.2.1), I set the intention to combine both a decolonial perspective and feedback from the community into a design. The outcome of this design would be including the perspectives of people positioned at the edges of the community. However, this was a broad and abstract intention. As I learned from my conversations, there are many ways to organize the community; those different representations also change who is positioned at the "edges." There are also many different ways to consider what it means to be included or involved.

While I attempted to weave together as many of the stories from my participants as I could, in practice I focused on those parts of my results that stuck with me the most. I started with the quotes I remembered, the phrases that stuck out to me, or the ideas that challenged or supported my own understanding of inclusion in the community (Appendix D contains an image of my initial design brainstorm with some of these starting places.) As much as possible, I brought these ideas back to the community to also keep my design goal relevant to the people I was designing for. In this section, I make transparent some of my design decisions and inspirations, explaining how they are connected to my results, feedback from the community, and literature.

5.1.1 Narrowing the Focus

Who to design for, what challenge the design will address, and how the design will operate – these are all decisions that I had to make to narrow my focus. Broadly, my challenge was "meaningful inclusion" and my target group was people that are positioned at the edges of the community.

On the scale of micro-communities

One of the first questions I asked myself is, how do you design for a community? The community itself operates as entity, but it is also made up of individuals. I had to consider on what scale I wanted to design. For instance, I considered designs for individuals – activities or tools that could be used by one person. I also thought about designs that would support interactions at an interpersonal or social level, things that might aid in the formation of relationships, that could facilitate certain types of communication or interactions between individuals. At a community level, I also imagined solutions that might benefit the formation or sustainability of communities, how they organize around each other, or how they define shared purpose. Since the focus of my thesis is on inclusion within a community, I decided not to focus on an individual level.

To understand a bit more what is happening in at an interpersonal or community level, I sought out people who organize communities. I talked with two community managers in charge of sustaining different university-wide communities: the Open Science Community, and the Delft AI Labs and Talent Program under the larger AI Initiative. One thing I learned from them is that larger communities are not active without a coordinator. Instead, smaller groups or micro-communities tend to have more initiative. This is aligned with how participants described momentum in the community as coming from individuals. Energy for initiatives was often from the bottom-up. Rather than trying to go against how the community already operates, I wanted instead to support it. For this reason, it did not make sense for me to try and implement a design from the top down that imposes a structure on the community. Instead, I decided to focus on these small groups, these micro-communities, as loci for creating change within the larger community.

Uplifting underrepresented perspectives

I wanted to ensure that this design somehow worked to uplift underrepresented perspectives. In Chapter 1, I talk about underrepresented perspectives as the various minorities, peoples, and cultures that have been historically excluded from academia. Indeed, many of the people I talked to from these types of social identities described difficulty relating to others, feeling different, or even feeling lonely in this community. Some talked about how they felt a need to represent their perspectives, whether that was through their research, speaking out about issues they see, or creating opportunities for others like them. However, I found it difficult to imagine a design at the scale of micro-communities that might uplift minorities in the community without also singling out and flattening their experiences. Similar to how treating the identity "person of color" as a binary overlooked the nuances of race, to design for "minorities" might overlook the complexity of intersectional identities and individual experiences. Whose perspective is underrepresented is contextual.

Instead, I tried to stay within the context of my research and decision to focus on micro-communities. From my results, I found that students, especially learners, are not as involved in networking events and various topic groups within the community. There were also participants

who felt as though their background or expertise was not well-represented within their faculty or lab group. Position and disciplinary or educational background are two aspects that I considered while making my design.

More specifically, I wanted to make sure to involve learners and young professionals. Focusing on young professionals also presented some opportunities. They are often present in these micro-communities. In my conversations, young professionals were the most likely to consider both professionals and learners in their own relationships or mapping of the community. Young professionals can act like nodes, connecting multiple levels of expertise. One of the community managers explained that they often try to encourage people with more time but less confidence to take on coordinating events or activities. When they shared this, I also thought of empowering young professionals to take on greater responsibility in the community. This is an idea that returns in some of parts of my design.

Many participants talked about how their educational or disciplinary background was an identity that influenced their approach to data science. My choice to incorporate disciplinary or educational background also came from the desire participants shared for more communication between disciplines. I also noticed a difference in values between people who have a different relationship to data. It was partially because of this difference that I wanted to involve the three levels of data science in my design (Section 4.2.1). The decision to incorporate this element also came from how two participants described the center of the data science community: as an intersection between the methods, the applied, and the philosophical.

Design Goal

From these considerations, I came up with an overarching design goal **to empower young professionals to act as bridges in forming or strengthening connections between micro-communities within the larger umbrella of TU Delft Data Science.**

This goal addresses the desire to have more cross-faculty events and greater communication between different disciplines. A few of the PhDs I talked with described themselves as "bridging" two fields, so it also builds on a strength that participants already recognized. It supports goals shared by the community managers I talked with as well. One explained that a challenge in their community is that very little is known about who is doing what. There is a lack of avenues to share this information between (or even within) faculties. An idea they had was to have faculty-level groups act as conduits of information. The other community manager discussed a desire to increase visibility and bring more of these micro-communities together under the larger umbrella of AI, Data & Digitalization at TU Delft. They explained that increasing visibility of these micro-communities could help connect more people to the group that meets their specific needs. Both of these managers suggested that having more visibility or connection points between these smaller groups could help with information sharing and bringing in new members.

After deciding that I wanted to focus on the increasing interactions between smaller groups within the larger data science community, I still felt stuck on *how* to accomplish this. I again turned to my community for ideas to make this focus actionable. I invited two of my friends (one bachelor's student and one PhD, both involved in data science) to help me brainstorm ideas. I prompted them with two questions: 1) how can we support micro-communities to learn about other micro-communities; and 2) how can we get these communities to interact or collaborate more? The outcomes from this session are in Appendix D.

Two ideas came out of this brainstorm and informed a more specific goal. First was a comment from the bachelor's student who explained that although they are interested in joining lab groups, they are often unsure what they could add to those groups. The second idea came from the PhD who felt that it was important to normalize collaborating on smaller projects, especially those related to the process of data science (not just the outcome).

These two ideas complimented comments from participants. Many researchers (especially professors) pointed out that the constant need to publish leads to competition and a lack of time. Focusing on other benefits of collaborating could help counteract this competition, which is also born out of the widespread culture of domination and superiority described by hooks (2003). Like other learners, this bachelor's student also felt unsure where they could contribute to the community (despite having a desire to do so).

Refined design goal: an activity to help people reflect on

1. What types of people or groups they might want to collaborate with
2. Ideas for collaborating all stages of research or projects (i.e., collaborating on things that won't necessarily lead to publications)
3. As a way to normalize working together on smaller projects and create greater exchange of resources and knowledge.

5.1.2 Collaboration as Relationship Building

In thinking about collaborating in all stages of research, I was reminded of the Declaration on Research Assessment (DORA) initiative mentioned in Franklin et al. (2023) that is challenging existing ways of assessing research focused on publications. One of their projects is to provide a set of resources and toolkits to reimagine research assessment.[1] In looking through these resources, I also thought back to how Streck (2021) considers the quality of relationships formed in the research process as a measure of validity. Why not also consider the quality of relationships as a measure through which to evaluate collaborations?

Supportive relationships

My design focuses on the role of supportive relationships as an essential part of meaningful inclusion. From the beginning of my project, I was inspired by the indigenous concept of *all my relations*, which speaks to the importance of being in good relation with people, creatures, and the land for our well-being.[2] Building ongoing relationships is also an essential aspect of decolonizing research (Datta, 2018). "Making the relationships rather than the agenda the priority of the meeting creates a space where participants can make genuine contributions" (Fish et al., 2022, p. 46).

My own results also highlight how relationships can impact both representation and redistribution. Access in the community often depends on who you know or on having knowledge of what resources are available. Close relationships were characterized by some in the community as also proactively providing more resources or opportunities. Many of the resources that were

[1] Project TARA (Tools to Advance Research Assessment) has published a collection of one-page documents with tips and suggestions for people in various academic positions. These toolkits also provided inspiration for the form that my design took.

[2] I was first introduced to this concept through a podcast named after it. All My Relations is hosted by Matika Wilbur of the Swinomish and Tulalip peoples and Adrienne Keene from the Cherokee Nation.

discussed by participants were also related to people – events, mentorship roles, lab groups or research groups, or even providing feedback for curriculum changes or regulations. An indication of supportive relationships was also the ability to exchange ideas or information – another form of redistribution. If my design goal is to increase the exchange of resources and knowledge through collaboration, strengthening relationships is a good place to start.

Feeling understood was the another aspect of supportive relationships that is related to representation. Participants felt support from those in the community who understood their work or experiences. One of the community managers shared a similar observation that smaller-communities organized around a purpose or need tend to attract people from a similar stage in their career because they have similar questions. This was also true in how participants were inclined to reflect with others in a similar position. However, lab groups are made up of multiple positions and many of the data scientists I talked with are also working interdisciplinarily. Returning to the quote from Fish et al. (2022), focusing on relationships rather than the agenda (or in this case, the goals of the collaboration) can facilitate understanding between different positions or disciplines.

Inspiration: Pod mapping

For my design itself, I was inspired by a concept that emerged from transformative justice work as a way to respond to harm and violence. Mingus (2023) developed an activity to help make the ideas of *community accountability* and *support* more actionable. Community is both an ambiguous and abstract idea – something I observed in my own research attempting to define it. Instead of placing responsibility on an abstract community, Mingus instead describes a "Pod" as the small group of people we can turn to within our own network or community. To intentionally build relationships in this way resists isolation, helping everyone become more resourced and supported (Mingus, 2023). This fits with the goals I defined. It also offers a potential way to counteract the loneliness that one participant described – sometimes just recognizing the ways we are already supported is enough to help us feel connected.

I liked this idea of pods, because it also fit with how I was imagining micro-communities. Pod mapping is a way to identify the people we can rely on for support and to fill certain needs or purposes (Mingus, 2023), just as micro-communities organize around specific needs or purposes. Mingus (2023) describes pod building as an ongoing process, starting with 1) defining the purpose for your pod; 2) identifying who already is or who you might want in that pod; and 3) reaching out to begin deepening those relationships. I adapted these steps to fit into an interactive activity that might be done together with a group.

5.2 Design: Reimagining Collaborative Community

My activity contains a one-page information sheet with some of the background I presented in this chapter, like the importance of focusing on relationships in collaboration, involving young professionals, and challenging existing ways of assessing collaborations. This is followed by a one-page activity that contains four steps.

5.2.1 An Interactive Activity

This activity is intended for labs, centers, research groups, or interest groups who are involved with data science research. It is an interactive reflection and all members of a group can (and

should) participate. Together, groups will 1) visualize how inclusive their group is of various disciplinary perspectives in data science; and 2) identify connections with other individuals, groups, or resources in the community that might bring more perspectives into their group. Data science research is an interdisciplinary endeavor, combining fundamental algorithms and mathematics, domain-specific data and theories, and societal motivations or implications underpinning the kinds of questions we ask. Including all of these perspectives can lead to new ideas, allow us to understand and better support others, and help us feel more connected to the larger data science community.

Below I list four potential benefits of doing this activity, each associated with a different step in the process.

1. It can be a way to build support within your own group. Research professionals feel supported by those who understand the challenges and significance of their work. Reflecting together on where group members are positioned in the larger landscape of data science is an invitation to learn more about people within the same group and how they understand their own work or expertise.

2. It can provide space to align goals and reflect on values in the group. Part of the activity involves discussing what kinds of collaborations or connections you might want to strengthen as a group. This discussion invites conversation about also what your group values and underlying motivations for your research.

3. It can help all members of your group become more connected and resourced. You might not know if your connection will help someone else. By making the existing relationships group members have transparent, it can increase accessibility of that connection for all members of the group.

4. It can provide ideas for how you or other members of your group might approach future collaborations from a perspective of relationship-building.

Step one: Map your group together

Mapping our existing group is the first step in identifying where we might benefit from strengthening existing relationships or forming new ones. In this step, group members each place themselves in the diagram shown in Figure 5.1. They can place themselves based on both their skills or expertise, and their area(s) of research. In this way, all group members can see how others position themselves within the data science landscape.

After each group member has been added to the diagram, they can share briefly why they position themselves in a particular place. This step is a form of recognition (as defined in Chapter 2), as they have an opportunity to reflect on their own position. Hopefully this also helps others in the group better represent them, as they can witness how each person would place themselves or define their own research and expertise.

Step two: Identify growth areas

The second step is a discussion step. The goal is for your group to collectively consider ways that your group might connect with more of the data science community and involve other perspectives. This step is similar to the first step of mapping pods: identifying a need or purpose that others can support with.

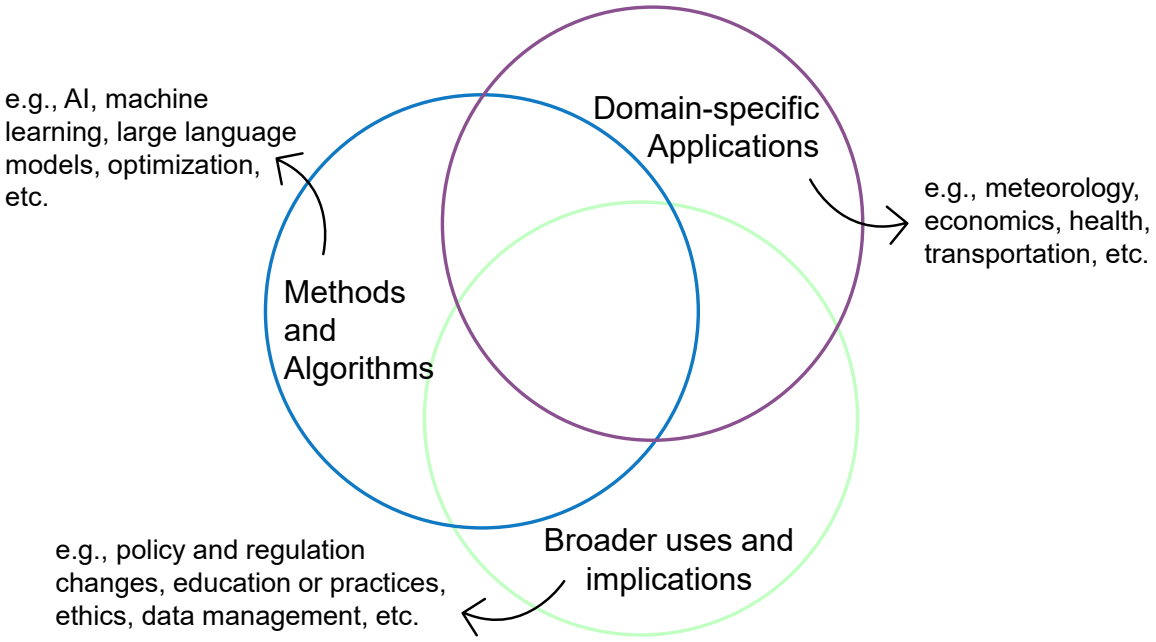


Figure 5.1: Diagram of three circles representing different levels of the data science field.

The diagram from step one is a starting place for this discussion. For instance, it might be very clear that all your group members are working on domain-specific applications, but none are directly involved in building or refining methods and algorithms. This might be a clear indication that your group could learn more from that part of the community. Sometimes it might be less clear what kinds of collaborations or connections your group might benefit from. In this case, I provide a list of potential "benefits" of collaborating.

Given more time, this list could be refined with more input from people in the community, so that it is based on things that worked for others. As it is, these ideas were taken from a few informal conversations, ideas about supportive relationships from my results, and from watching my own lab group interact and collaborate cross-disciplinarily. It was also inspired ideas published in a toolkit for Rethinking Research Assessment.[3]

What are some benefits of collaborating on projects that are not tied directly to research outcomes?

- Strengthen relationships that can lead to future collaborations
- Collaborate on pre-processing steps that can save time in the long run (e.g., methods for cleaning a particular type of data)
- Give or receive support during challenging times
- Provide richer cross-disciplinary experiences for students
- Contribute to the research community or society in different ways (e.g., sharing resources, providing information for a public audience)
- Increase group visibility
- Etc.

[3]Hatch, A and R. Schmidt. (2020) Rethinking Research Assessment: Ideas for Action. DORA. http://sfdora.org/wp-content/uploads/2020/11/DORA_IdeasForAction.pdf [Accessed: 18 Oct 2023]

Step three: Start from existing connections

One of the concepts from pod mapping is to start from your existing connections – it is easier to strengthen an existing relationship than to build one from scratch (Mingus, 2023). The third step is about reflecting on our own connections and sharing that information with your lab group. Ideally, each person can try to think of individuals, groups, or resources they know about that might fill the growth area identified in step two.

An important consideration for this step is to especially encourage students in the group to recognize their own interconnectedness. In interviews, this group was the least likely to draw actual relationships with specific people. However, they also have a lot of connections, especially to other students. This could be a way to help students, or other individuals, realize that they do have a strong support network already.

Of course, it might also happen that an individual in the group struggles to think of any connections, especially within the data science community. I included a small prompt in the design to try and address this. It prompts the group to consider how they might support or become a more significant part of this person's network. One participant shared that they did not feel supported by their own lab group; this activity could also make the group aware that there are people within their ranks that do not feel connected or supported.

The other key aspect of this step is seeing the connections that others in the group might have. I found in my own research trying to reach out to potential participants that I was more likely to get a response when I was introduced to someone through another connection. Sharing the names of people with whom I have an established relationship is a way of offering to connect others with my network. This is a form of redistribution that supports inclusion. Sharing openly and making data or resources accessible were values that many participants held. This step expands on that value by also encouraging us to consider sharing relationships or connections to people as well.

Step four: Ideas to strengthen relationships

While sometimes recognizing how we are connected is enough to "strengthen" a relationship, I also wanted to provide actionable steps for groups to take. This desire came from my co-creation sessions (Appendix A), where we reflected on how so much conversation around inclusion remains just that – conversation. Below is the list of ideas that I include in the design for ways to strengthen relationships. As with the list of benefits of collaborating, these ideas are a combination of informal conversations, various resources I have already mentioned, results, and my own experiences.

- Ask if someone who has helped you before would be willing to lend similar support to other members of your group. This is a way of asking for consent, which Mingus (2023) emphasizes is an important step in the pod mapping process.
- Invite someone with similar interests to attend an event together. This was inspired by the "bring a friend" policy one of the community managers has, and an idea from Mingus (2023) to attend workshops together for mutual learning.
- Proactively share sources, information, expertise, etc. with those who might benefit from it
- Collaborate on projects focused on pre-processing steps (e.g., data management systems,

efficiently cleaning data sets, etc.)

- Discuss your values in research with students, supervisors, or existing collaborators – especially those in a different position. This is from my own results and learning that most reflection opportunities are informal with peers or with supervisors. Through this suggestion, I also hope to invite people to consider reflecting with others whom they might not normally.
- Share (formally or informally, internally or publicly) about setbacks, challenges, and things that don't work in your research process
- Involve others (especially young professionals) in planning or decision-making processes. Inspired by results, this is a way to both empower young professionals and increase transparency in the community.
- Get involved with educational projects or courses around campus by providing data for assignments, real-world examples, or having them reproduce a small part of your research, etc. Multiple PhDs discussed how they either are or could imagine using their research to support courses. This is another way of also involving learners more actively in the community.
- Don't know anyone who fits the expertise you are missing? Make a job posting on your website highlighting the kind of student or collaborator you want to work with!

Putting this activity into practice

Facilitation:
While the intention is for this activity to be self-facilitated, in practice it will likely require a designated facilitator to guide the conversations associated with each step. There are some important considerations for how to facilitate this type of activity that I did not include in my first design iteration. Things like: making sure to establish a space in which all people share and contribute, or establishing "ground rules" about listening, etc. Another aspect of facilitation is knowing when to give people more space to talk, and when to let someone else take the floor. Especially for some of the conversation and reflection points like in step two, this might be very important.

I also do not specify who should facilitate this activity. In most lab groups, I would expect that a PI or professor would initiate this activity, and so I wrote my activity with this position in mind. They have greater permanence in a group than students. However, I would also encourage a young professional to lead this, to upset the existing hierarchy in academia and promote greater initiative and responsibility from this group.

Timing: *Estimated 10-30 minutes*
Time was a major limitation that participants talked about – it is a major limitation I experienced as well. One of my initial hesitations with designing an activity is that activities take time to organize and facilitate. A hope I have for this activity is that it does not demand too much time. I did a test of this activity with my 4 people and it took approximately 20 minutes. With a larger group or with extended conversations, it would likely take longer. However, if time is a constraint, this could be 10-20 minute activity appended to an existing meeting. It is also possible to put time constraints on each step of the activity.

Activity size: *2-10 people*
Based on my own experience facilitating interactive activities, I believe this activity would be most appropriate for 2-10 people. When groups begin getting much larger than 10 people, it becomes difficult (and takes a lot more time) for everyone to share. While it might be possible to do this with more than 10 people, it would require facilitation by someone who is comfortable with guiding discussions and managing the logistics that accompany having a large group. For instance, considering how to make sure everyone can see the diagram and reach it to draw their connections. Or giving time for everyone to do each step independently before coming back to share, so as not to have a few people overpower and dominate what people share.

This activity will also be more successful with at least two people. This is because some of the value of this activity comes from it being interactive. It is an opportunity to 1) learn about how others position themselves, 2) discover either shared connections or connections through others, and 3) negotiate together what types of benefits or collaborations might be important to you. In theory, this could be an individual activity with different intended outcomes. A person could independently consider where they might place themselves in the data science landscape and reflect on where most of their connections fall. This can still provide value by supporting reflection and recognition of an individuals position. It also more closely follows the intentions of pod-mapping, in which we identify the specific people whom we can rely on for support in different areas of our life.

Potential Adaptations:
These types of activities are typically easiest to do in a physical setting. However, it is not always easy for people to gather in one place. This activity can be run via an online meeting, as long as there is a way to have a shared diagram that everyone can interact with.

This activity could also be adapted to work asynchronously. Instead of everyone being present and doing all steps together, people could do the steps independently but share them with others. For example, this activity could be converted to a "prompt" for participants to share their own maps (steps one and three) on a group communication channel. This might also be a way to adapt the activity for larger groups.

5.3 Validation and Refinement

In describing my design, I give examples of how different elements are informed by my results or resources and literature. However, I did not have the time to properly test and validate this design.

This said, I can draw on personal experience to at least justify the potential impact that focusing on relationships can have. In the last few months of this project, as I became more and more constrained by time, the easiest people to ask for help from were the people directly around me. Other science communication students in our study room, my teammates whom I already felt comfortable and close to, and my housemates. These are the people who 1) checked in on me regularly, and were then able to offer help when I needed it; 2) who I felt comfortable being a wreck around because we had already established a relationship where we shared vulnerable moments or difficult challenges; and 3) almost all of them I had received help from before or given help to in the past. For this reason, I believe it can have an impact to focus on building relationships through these last three steps. Regular communication, space to be vulnerable or share challenges, and an established practice of both offering and receiving help in reciprocal

ways.

Although I can imagine future uses and potential limitations to my design, these are based on speculation. Given more time, I would have sought out more feedback on my design and maybe even revised it.

5.3.1 A Generalizable Activity

While this version of th activity was inspired by the interdisciplinary nature of the data science community, it could be used to consider other types of inclusivity as well. For instance, the three circles could be replaced with circles representing different regions of the world to consider whether your group is inclusive of global perspectives. It could be redrawn to represent differing social identities (e.g., gender, race/ethnicity, culture, etc.), work or educational experiences, institutions, or even more abstract concepts such as values or priorities. For those interested in responsible research innovation, the circles could also represent the triple helix of government, academia, and private industry.

I also considered how this activity might be useful for those in more "managing" roles within communities. One of intentions of this activity is that people might place themselves somewhere that others would not think to place them. This aspect is about reconciling recognition with representation, and seeing others how they want to be seen, not how we assume they are. However, it might be interesting for someone in a position where they are connected to many people (such as a community manager) to think still reflect about where most of their connections are. It could also be a way to reflect on which connections are more "active" or stronger. So rather than focusing on *where* people might be positioned, instead reflecting more on the quality of relationships themselves. The activity then becomes more similar to the concept of pod mapping. However, perhaps having a diagram in which to position people still encourages some reflection on the composition of close connections. For instance, considering if you tend to invite more people from a certain position to events. Or whether you primarily offer resources or ask for help from people who share a similar cultural background.

A final adaptation I considered for this activity is making it a tool to facilitate the beginning of a collaboration. When starting a new project, researchers don't necessarily know what types of networks, people, or resources their collaborators might have available to support the project. Learning about these is usually part of negotiating a new collaboration. Especially if the collaboration is between two people in different lab groups, universities, or even just from differing disciplines, there might be a lot of connections that are not known. I imagine this design could be adapted as a way to both align goals and learn about resources. Instead of starting with the three circles representing levels of data science, collaborators could start by drawing their own overlapping circles. The first step would then be to define together what will be necessary for their joint project. They could still think about this in terms of methods, background knowledge, and overarching goals or societal implications; or they could define something different entirely. Perhaps even just having the different disciplines cross-disciplinary collaborators come from would be interesting to explore. The rest of the steps could then be about determining what people or resources each researcher can bring to fill these different areas (for instance, maybe a PI has a PhD in their group who is an expert in a particular method). Such a tool could suggest a way to start a collaboration from a relationship, rather than from an offer of resources.

5.3.2 Feedback and Design Limitations

One of the challenges that a community manager described is trying to balance the needs of a diverse audience. When organizing events, they try to oscillate between broad and niche topics to keep the interest of as many people as possible. I experienced a similar difficulty trying to design for the data science community. I wanted to on one hand capture what is meaningful for me (based on my results and my own experience), while also creating something that is valuable to others (such as the lovely community managers I talked with). While I have created something I find interesting, I am less sure about how valuable it will be for others.

My biggest disappointment in my process was not having the time to properly get feedback on and refine my design. Although I was able to involve people in much of my design process, I was not able to get feedback from anyone directly involved in the data science community.

That said, I intend to send a version of the design to both community managers, along with a few former professors for whom I anticipate the ideas I present might be valuable. This is also my commitment to share the outcomes of my research and redistribute the knowledge I have learned.

Within my design itself, it is lacking detailed facilitation notes. This type of refinement is easiest to do with proper testing, so you can learn what parts of the activity a group finds simple or confusing to perform. I also would have liked to get addtional feedback the list in step four, and ideas for collaborations from step two.

Finally, I made a decision to not make a design that specifically considers the way minority perspectives are underrepresented in the community. However, this decision is also a limitation. This activity is designed for people who are already in the community; it is designed for the people already in the room. It doesn't address the bigger issue of bringing in and supporting the perspectives that are not there at all.

So much of the rhetoric around diversity and inclusion is idealistic, with very little tangible changes. This is often most felt by the people who benefit the least from the existing system – by the people of color who still feel such a strong obligation to represent their people. My desire to do something great with this design was in constant tension with the pressure to finish on time. This tension is also result of writing this thesis and doing this project within a university. At the end of the day, I was still embedded in a system that operates on the myth that we must constantly be producing. "Good things happen when you take your time."[4] Unfortunately, time was my biggest constraint.

[4]This quote is from visual storyteller Brit Hensel (from the Cherokee Nation) describing her artistic journey on the podcast All My Relations Podcast: Telling True Stories in a Good Way (March 16, 2023). She talked about how she spent two weeks in a community rather than her originally planned two days, and ended up with beautiful relationships as a result.

6 | Discussion

During my exploratory phase, I learned an incredible amount about the community. However, I also learned from the community. Many of the ideas I write about in this thesis originated from participants. Their values and how they defined them guided my interpretations of literature and my approach to data analysis, design, and writing this thesis. In the first section of this chapter, I revisit my discussion of different results and bring them together. I highlight strengths of the community and the ways they are already contributing to meaningful inclusion and disrupting power systems. Beeman-Cadwallader et al. (2012) emphasize that researchers should not ignore the strength of indigenous communities they work with by focusing on problems; in the same way, I also do not want to focus on limitations within the TU Delft data science community. In the second section, I take a step back and think about how my understanding of meaningful inclusion has evolved and what that means for this community and beyond. Finally, in the third section I reflect on limitations of my process, some of my own biases and the larger impact of this work.

6.1 Learning from the Community

The data science community at TU Delft is not a monolith; the people I talked with had different research, different goals, different ways of interacting with the community and others in it. This was clear from how many ways the community itself could be represented (Section 4.2). The identities of people within the community are also both complex and fluid. We bring different identities into different parts of our work or life (Section 4.4.2) and don't always fit into the labels for identity that are placed on us (Section 4.1.1). Boyd (2021) explains that appreciating the complexity of our own interdependent social identities lets us use those identities to confront structures and systems of power. They write about using an intersectional framework to approach data science, as a way to counteract the way big data and data colonialism separates people from the data that belongs to or describes them (Benjamin, 2019; Thatcher et al., 2016).

Appreciating our complex identities is also a way to challenge the separation that can occur within the research *community*, not just within data science research. The disciplinary structure of academia is rooted in imperialism and creates a distance between researchers, absolving them from being responsible for what happens outside their field of expertise (Smith, 1999). I was inspired by how many people I talked to were working on interdisciplinary or transdisciplinary projects. Data scientists at TU Delft are challenging this disciplinary structure, which perhaps also explains why they feel such a strong responsibility over both their research and the impact of their research (Section 4.3.1). However, many professors bridging disciplines felt as though they do not fit in with their faculties or were not understood by other researchers. They still felt a pressure to adapt to the expectations of journals and the structure within the university (Section 4.7.4). We still have expectations about who is a data scientist and how they fit into existing structures. For instance, the EEMCS faculty which hosts computer science (and soon a data science master's program) is often associated with the data science community. However, I learned in that researchers within this faculty do not see themselves as central to data science, but rather on the periphery. The data science community does not seem to be a system with

centripetal movement as described by Fish et al. (2022). The community is decentralized and organized from the bottom-up (without one person, group, or initiative carrying more than the rest). Data scientists are not only the mathematicians and software engineers in EEMCS; they are also working on policy, engineering, fundamental research, and design projects across the university.

In my results, I define "data scientist" in a way that supported my analysis of this community and allowed me to talk to a broad range of people. "There are social choices in what we choose to measure" and these choices can be anchored in our existing expectations about the world (Franklin et al., 2023, p. 4). How we define certain groups is important for data science research, but it is also important for how we navigate community. It can affect who is invited to events (Section 4.6.1) or who we go to for support (Section 4.5.1). Institutionalized expectations can ascribe difference to certain groups or fail to acknowledge their distinctiveness, both of which prevent full interaction (Rosa, 2017). This is evident in the expectations people have about the types of research different faculties are working on. It also shows up in the academic hierarchy and how students are seen as distinct from professionals. Such a distinction can prevent them from being seen as valuable contributors to the community, or impact how they perceive their own capabilities (Section 4.4.4). Data scientists in this community value reflecting on and transparently communicating their research decisions. These same values can also guide how we approach interacting with a community. How do our biases impact what we expect from people with different identities or positions? How are we making decisions about who to organize events for, who we interact with, what opportunities we pursue or ignore?

There are a lot of opportunities and resources available at TU Delft, yet many participants still desired more (or different kinds of) opportunities (Section 4.7.3). They explained the data science community could benefit from more reflection opportunities, more cross-faculty interactions, and more events aimed at understanding the research of others. What we value shapes the opportunities that we desire; the opportunities that are offered can also shape or reinforce shared values in the community. The community manager I talked to from the Open Science community explained that having institutional alignment with your goals makes it much easier to get things done. Kirkness and Barnhardt (2001) similarly writes that universities are set up to integrate individuals into the existing culture and become aligned towards reproducing similar goals. However, I also saw examples of the reverse – the ways individuals in the data science community are shaping their own purposes and goals despite the university culture. This made me realize that individuals (myself included) can play a much bigger role in shaping a community than I previously believed.

Rosa (2017) highlights that symbolic voice plays is an important part of participating as equals, but that social and cultural hierarchies can undermine our ability to develop voice. This seemed evident in how students felt less agency to initiate reflection opportunities or did not see themselves as experts able to give back to those higher than them in the hierarchy. However, Kirkness and Barnhardt (2001) explain that reciprocal teaching and learning between students and professors is necessary to make higher education spaces more inclusive to First Nations students. Reciprocity allows those already in the university to adapt to and understand the culture of those coming to the university. Reciprocity in decolonizing literature also has a similar meaning to how participants described being open-minded to the perspectives and experiences of others (Section 4.3.1).

The diversity of values participants mentioned highlights that people in this community have different priorities in their research, work, and studies. Having different priorities is important to

challenge issues of representation that occur in research, especially research concerning social demographics or populations (Franklin et al., 2023). However, there was a difference between how participants from EEMCS and TPM talked about values (Section 4.3.2). People affiliated from TPM demonstrated and discussed more values, especially those related to the social environment, like respect and empathy. Talking about what we value in a work environment is important, because it is a way of refelecting on the cultural expectations that might be embedded in how the community operates. For instance, there is an expectation at TU Delft that people will ask for or seek out the resources they need (Section 4.6.1). From this expectation, it seems the TU Delft values being proactive – but that was not a priority mentioned by participants. Open discussion about values can help us bring into question aspects of a the community we might take for granted. This is a way to understand the lasting impacts of colonialism, to see how our university might still unintentionally exclude people.

One participant talked about how the responsibility for changing a system always lies with people who are victims of the system. In my conversations, I also felt that the responsibility to challenge dominant perspectives was felt most strongly by people who identified as a minority (Section 4.4.2). Similarly, the students who responded to my interest form and were willing to participate in my research also all identified as part of a minority group (Appendix B). Sultana (2018) highlights why it important to shift this responsibility to represent to fall instead on those who benefit most from this system. Sultana noticed that activists and scholars visibly trying to fight social injustices often become the targets of attack and abuse from far-right or white-supremacist groups. They call on more academics to join collective action and be visible in their stance to protect the same people from always being targeted. Making Standing in solidarity with others is a form of representation by reflected back to us (Smith, 1999). Decolonization within an academic context starts with recognizing that we have to take an active role in making a more inclusive environment. Part of this recognition is also acknowledging the work people are doing to change the community, especially when it does not directly impact us. This is how I interpret uplifting underrepresented perspectives to solidify both place and identity (Datta, 2018).

6.2 The Power of Recognition

The data science community is a space where researchers can pursue their interests and values. That freedom is why some people stay in academia, because they can work on projects they believe are important. But this freedom, this agency, also emphasizes the importance of reflecting on our position. Values, culture, and perspectives are also learned from those around us. Many of the values described by participants were shared; some of these values were learned while others were believed by participants previously. Yet it can be difficult to determine how much of our perspective we define ourselves, and how much is shaped by the culture around us. And while some people in this community feel the pressure to adapt to the system very poignantly, others did not even mention it. How we represent ourselves internally and how we are represented externally can become conflated.

Recognition is a way to counteract this, to build awareness of how we fit into a system. However, sometimes our awareness of how we don't fit into the system can make us feel even more marginalized. One participant explained, "the pressure of being different, it's immense." Yet while some participants saw their differences as a challenge; others saw it as a strength. Recognition is a powerful tool because it can change our own feelings of marginalization and allow us to take greater ownership over our own participation. It can support us to better understand

our role and responsibility in that system – and also empower us to change it.

This is the reason I focus on recognition in my design. It is also the simplicity behind the concept of pods; through making explicit who is part of our "community", we can hold individual people accountable to create change. It is difficult to make an abstract community more inclusive. Instead, we can think about our individual role in and responsibility to create a more inclusive community from the bottom up – starting with how we see ourselves.

6.2.1 Our Role in a Changing Environment

Data science is evolving at an incredible speed. The data science community at TU Delft is also undergoing major changes. The launch for this academic year (2023-24) was all about AI. Between when I started this project and now, the information I could find from searching "TU Delft Data Science" online is changed. The university-wide AI, Data & Digitalization initiative is trying to become more visible and link together a bigger part of the research community working on these topics.

It is important to reflect on our position, because it shifts as we move through and interact with this rapidly changing environment. For instance, the role of "expert" is not a fixed position. New tools, methods, and data are constantly becoming available. Thinking of an expert as someone who possesses tools to do their research or familiarity with data means that people can become less expert as the community changes. Rather than holding these skills as a measure of expertise, perhaps it is more valuable to consider experts as the people who understand and are able to communicate the implications and importance of what they are doing transparently. Likewise, the influence of institutions within data science can fluctuate. Some participants feel more impacted by the regulations imposed by government, while others are more constrained by the data available from companies. Our relationship to these institutions is not fixed; it changes based on what the institutions contribute to data science and based on our needs.

How we see ourselves also changes in relation to this dynamic system. This was clear from how participants talked about the data community, switching between different representations throughout our conversation. They would also switch between different ways of placing themselves within the community – sometimes based on position, sometimes based on expertise, other times based on their type of research or their disciplinary background. In such a community, there is no fixed "center" around which everything revolves. The community itself is a dynamic system.

The myth of fixed positions in a fixed system benefits the same people who already have the most power in a system. It is a way of concentrating privilege and encouraging conformity because we do not imagine something different (much like what happened in the Aerospace faculty described in Section 1.1.2). Fish et al. (2022) describe "brokers" as people in the center of a community who can bring those in the periphery towards central participation. This places the burden on new participants to adapt to an existing community structure by starting on the outside and moving in. Even the concept of a "center" is a Western spatial term meant to describe our orientation around a system of power (Smith, 1999). To reimagining ourselves in a different position that is de-centered opens up space to have different power structures and disrupt existing inequities (Fish et al., 2022). Recognition is a decolonizing practice in so much as reflecting on our position allows us to redefine a new "center" to the community. This is one of the desired outcomes of the activity designed in Chapter 5. The activity invites groups to establish their own "centers" around what their group values, while also considering how they

might involve different disciplinary perspectives.

6.2.2 Our Relational Responsibility

"Decolonization is an on-going process of becoming, unlearning, and relearning regarding who we are as a researcher and educator, and taking responsibility for participants" (Datta, 2018, p. 2). Datta writes about doing research with communities, however this quote is also relevant for research communities themselves. Becoming, unlearning, and relearning are about recognizing our role. Taking responsibility for research participants is about relationships. How do we take responsibility for our connections within a research community – for our supervisors, mentors, mentees, peers, and collaborators?

More than half the people I talked with valued responsibility. They described responsibility in many ways, but all of their definitions were related to taking action. Responsibility in relationships can be understood as the ongoing obligation to show support to those around us. Returning to the idea of reciprocity (as explained in Section 2.1.2), relationships are not just about what we can get from others, but also what we can give to them.

Building a research community requires investing time into a community beyond just building rapport with others (Falcón, 2016). I propose in Chapter 5 that we can strengthen connections within our research communities by putting time into (and prioritizing) our relationships. Participants explained to me that connections in the community often occur through a snowball effect, where one person introduces someone to another person. This is a way through which redistribution occurs in the community. It can also be an opportunity to represent more perspectives in the community as well. We learn from people who have different opinions and perspectives than us. Yet, people in this community still had a tendency to seek out others in the community with similar experiences to reflect or get support from. Another aspect of our relational responsibility is to also seek out diverse connections. This echos the responsibility felt by many participants to bridge two disciplines. It also echos the how one participant described their responsibility to increase representation when they have the opportunity to hire others. When we are in a position to connect people from different disciplines (or cultures, identities or perspectives), then we should be bringing them to the table.

Both my results and the activity I designed are relevant to a very specific context. How people in the data science community at TU Delft feel about inclusion here is not the same as others might feel elsewhere. The part of this thesis that *is* relevant to others is the way these participants helped me describe a form of inclusion that is actionable. The experiences of this community helped me reframe inclusion as an individual responsibility to show care in our own relationships. This perspective is relevant to all kinds of contexts.

6.3 Reflecting on this Project

6.3.1 Limitations

It is challenging to accurately represent a group as large as the TU Delft data science community with only a handful of responses. My choice to use a time-consuming method meant that I could not cover the experiences of people in the community broadly. I did not get an opportunity to learn about the experiences of people in other faculties or in non-research or study roles (e.g. data stewards, staff, or administrative positions). I also did not include any perspectives from

outside academia, despite the TU Delft having numerous connections to both industry and government. Instead, I chose to look more deeply at fewer perspectives. However, my depth of analysis was also limited by time; there is no clear stopping point when analyzing exploratory interviews.

There were limitations within my scope as well. The timing of my data collection at a busy time of year prevented a lot of people from participating. While I expected professors would have a low participation rate, this resulted in fewer students participating as well. I was not able to talk with people from all positions; instead, master's students and PhDs are overrepresented in my sample. My attempt to scope for identity also complicated analysis. As I explain in Section 4.1, by treating the identity "person of color" as a binary, I simplified the nuances of race and appearance in a way that made it challenging to properly assess the influence of this aspect of identity. A secondary consequence of this is that I found it difficult to also design an intervention that particularly promotes inclusion of racial diversity.

I also introduced sampling bias into my project through my framing. I realize now, after my own understanding of what it means to be a data scientist has expanded, that I was not clear enough about who I was including in this community. It is possible I got fewer responses from students, especially bachelor's students, because they didn't consider themselves data scientists. This unclear framing also meant that a few professors felt they were not the right target group.

Framing my project to be about inclusion also likely biased my sample towards people who already felt that diversity and inclusion are important. This was apparent in how professors responded to my emails and suggested others I could talk with. It was also apparent in the interest form I sent to students; the three students who took the time to fill out this form all identified as belonging to a minority group (Appendix B). This said, a sample biased to consider diversity as important is potentially even more valuable for me; these are people who may have already considered challenges or visions for how to bring more perspectives into the community.

6.3.2 Bias and Privilege

I was posed a question during my research: why are you the right person to tell this story?[1] Just as many of my participants talked about bridging multiple interests, I also see myself as a bridge. Growing up in a multicultural household, living and working in multiple countries, I have developed a language to describe the describe the feeling of liminality. I have learned to sit with the complexity of human identity, because my own identity is nuanced complex. At the same time, I am in a place of privilege. As some of my participants also described, I can move through society without my identities impacting my ability to live or work. I am at a well-respected university, which gives me a platform to do this work. I am positioned at an intersection between these two perspectives.

Yet these perspectives and the way I personally understand and value inclusion also biased this story. I chose literature to use in my framework that reinforced my own worldview. In analyzing the transcripts of others, I looked for their comments that also aligned with this framework. I did not include the parts of our conversations I considered irrelevant. Everything that I wrote about as important is filtered through my own values of what I believe is important. Ultimately, what I have written here may be more a reflection of my own perspectives than it is a representation my participants' views. My participants views, too, are also only representative of a small part

[1] This question came from an episode of the podcast *All My Relations*.

of the community. So although I write and ask about the data science community, this project is not a perfect model of reality. It should be treated more as a story; whether fictional or real, it can still provoke thought in readers.

At the end of the day, I am also still writing this within a system and adapting my story to fit the structure of a thesis. Throughout this process, I constantly questioned my own words and re-wrote sentences to try and tell the reader what a person said, rather than what I found. Every time I included a fragment of a quote, I wondered whether it was fair of me to take these words out of context, whether I was still being true to the meaning behind the participant's words. I was constantly challenging myself to not fall into habits of writing passively, to not dehumanize the stories that were shared with me. This is also a reminder to readers to remember that behind all of my work, there are real people with real experiences.

There is nothing new that I discovered; all of the things that I wrote about were already known by someone at some point. It was a privilege (and a joy) to learn so much from so many people. The novelty of this work is not developing a new activity concept or creating a new framework. The novelty of this work comes from the ways that I connected existing ideas together, and brought more people into the movement to reimagine, transform, and decolonize academia.

6.3.3 Centering Racial Diversity

In Section 1.1.1, I question of whether it is appropriate for me to call this thesis "decolonizing" at all. I fear that by using this terminology, I am appropriating misrepresenting the goals and values of a movement meant to center the people who have been most oppressed by colonization. Despite having an intention to center race in my own considerations of diversity, I instead focused my design and much of my results diversity in position or expertise. Instead of building on the existing efforts of others to decolonize academic spaces, I took their ideas and applied them to a university that predominantly serves white people coming from European or other Western backgrounds.

However, this does not mean that my work is not valuable or valid. Although the activity I designed doesn't directly address or reach marginalized groups, it is still a place to start. If those groups which are already in a place of privilege find it difficult to share, collaborate, or include each other, how can we expect that they will be able to do the same with those who are not coming from a place of privilege?

I still strongly believe that university-wide conversations about diversity at TU Delft need to center race in a way that they have not — in a way that even I was not able to. Speaking on behalf of less privileged individuals can deepen or maintain the oppression experienced by the group that is spoken for (Chin et al., 2022). This worry led me away from focusing heavily on minority identities, people of color, race, or even gender. Yet I also realized through this project how stepping away from representing those who are less privileged than us can create an even bigger burden on those communities. Despite that this is a conversation that needs to happen. The burden to create change is too often placed on minorities; and because they are a minority, they are even more overworked. This idea was shared by a participant and by one of the community managers I talked with, who explained that the same people often get asked over and over to represent the "other" group they belong to.

My hope is that writing from a decolonial perspective, I am able to inspire others to bring these ideas into other academic movements, such as the open science movement or transdisciplianry

research. Collins writes that "Black feminist thought is neither the intellectual property of Black women nor should it be. Rather, seeking out points of convergence with similar projects potentially enriches all parties involved, and certainly exposes how systems of domination work similarly across intersecting power relations" (2016, p. 138).

6.3.4 Impact through Community

The biggest impact of this work comes from my journey and the way that I have built my own research community. By making my process collaborative and sharing formally and informally about my own struggles and realizations, I was able to plant a seed in others. I challenged colleagues, friends, supervisors, and family to reconsider what inclusion means and to reflect on their own assumptions, biases, and interactions with an academic system. Through these conversations, I have built a supportive community around me that understands the value of this work and efforts much larger than me to decolonize research and education.

I also saw this impact also in my interviews. One participant explained that from our conversation, "I realized that I feel supported." Another participant described how "this conversation, I think it helps me with, you know, seeing more than what I'm seeing." Or one person, who told me, "I feel like you understand me." I held onto these quotes, because they were a reminder to me that I was able to touch others in a meaningful way. They also highlight the validity of my work, based on how Streck (2021) describes the quality of relationships we form with participants as a measure of validity.

While I cannot predict what kind of impact this work will have on others, I hope that people from all walks of life – academic and non-academic, students and professors, those who feel privileged and those who feel like they don't belong – can find something in this thesis to relate to. Although the results from this research are unique to the TU Delft, I believe that others can still learn from my process.

7 | Conclusion

If we consider meaningful inclusion as a combination of representation, recognition, and redistribution, the data science community at TU Delft is an environment in which these things can (and in some instances already are) happening. The data science community, at least the people I talked to, understand the importance of redistribution. They expanded my own understanding of what redistribution might look like at an interpersonal and structural level. For instance, communicating in a simple and clear way to increase accessibility, or sharing decisions transparently to give people insight and access into processes. These values that came from the community stuck with me, and even influenced aspects of my design.

Yet, while these values were felt to be shared in the community, the TU Delft as an institution does not always operate in accordance with them. Knowledge of resources can be difficult to find unless you know what you are searching for; ask any student about the TU Delft website, and they will tell you it is easier to google search for the information you need than to find it through the website itself.[1] Bureaucratic and disciplinary structures limit the knowledge that is shared between faculties and positions. In short, position does impact how people in the data science community access resources and form relationships. So while the community has good intentions to redistribute resources, aspects of the system need to change in order to do so.

Many participants similarly acknowledged the importance of increasing diversity and representation, but very few were able to name concrete actions that might allow for greater representation in the community. At the same time, I learned from participants about numerous small actions that they are already taking which can support representation. Relationships characterized by reciprocal reflection and learning provide support for and opportunities to understand the perspectives of others. Cross-faculty collaborations challenge a disciplinary frame of operating and create space for researchers to approach their work from many perspectives. Pursuing research that is aligned with their own values is a way for researchers to express and represent their own perspectives.

However, like with redistribution, aspects of university culture limit our ability to take these actions. Time and competition can discourage collaborations or sharing of resources and make it more challenging to teach, share or otherwise act in accordance with our own values. So while people in the community are able to express their perspectives, they do not always feel as though they can. In part, because the wider culture of academia does not celebrate or reward these types of small actions. Indeed, the rhetoric around diversity as a "target" or something to "increase" undermines the way that inclusion, like decolonization, demands ongoing work. If we can recognize how meaningful inclusion happens at an interpersonal level, maybe more people in the community will feel empowered to do this work.

[1] I am referring to the TU Delft main website: https://www.tudelft.nl/en/. I have often discussed with my friends about the difficulty of finding things on the website and have even been told to google search information rather than click through the website tree. However, I have a specific example of how this limits access to resources. I learned from one of my co-creation sessions that TU Delft offers financial support to students who need to prolong their studies for unexpected reasons. I looked around a bit for this information, but could not find it. Multiple months later, my friend mentioned a similar fund, but this time by name: the RFP fund. I was able to find it immediately through a google search.

Asking whether a *community* is inclusive makes it difficult to see our individual responsibility to take action. Instead, we wait for strategies from the university or for committees on diversity, equity and inclusion to make tangible changes. Support from the top and alignment with university-wide goals does make it easier to take action (since individuals are not pressured to adapt to the dominant culture and can receive funding or other resources). At the same time, much of the momentum in the TU Delft data science community is generated by individuals. Recognition challenges us to reconsider our own role in a community; it is a way to empower more individuals to generate momentum and create opportunities for redistribution and representation. Some of the people I talked with in this community have already recognized this responsibility. These are the people who are speaking out when they see problematic behaviors in their peers. These are the people who feel a strong responsibility to represent groups that are overlooked by most research. These are the people proactively sharing information about resources and events, rather than waiting for others to come looking for it. These are the people trying to bring more diverse backgrounds into their own circle of connections.

I believe this shift in mindset from seeing inclusion as a community issue to understanding it as an individual responsibility is the most impactful way for the data science community to uplift the perspectives of those who are positioned at the edges. Our position within a community is dynamic, changing based on our relationships and interactions with people and processes. Someone who is positioned at the edge of, for instance, the applied research community might feel central in data science education circles. By focusing on those people directly around us, we can see how our position changes in different circles. We can identify those circles in which maybe we have more privilege, more ability to act. Understanding our own privilege in different contexts is the first step to using that privilege to uplift others. As one participant explained, "your privilege puts you below someone. You have that weight on your shoulders, and you have to use it."

Taken in its entirety, I cannot conclude whether the TU Delft data science community is inclusive of diverse perspectives, especially those from peoples underrepresented within academic spaces. However, individuals within this community are creating parity in participation. Some have recognized their own responsibility to use their privilege and uplift others when they can. But this is an ongoing effort. And as an institution, the TU Delft can do more to recognize, celebrate, and reward the work that is being done.

A New Perspective

I spent a lot of time considering whether my own research embodied both decolonial practices and the type of meaningful inclusion that I wrote about. This thesis is by no means a perfect example. But I have tried, where possible, to represent the stories of people in the community as accurately as possible. My commitment to share knowledge, resources, and insights with others has extended into my own life and personal relationships. I understand better how I fit in the research community at TU Delft and the impact that I can have one those around me. Yet I also experienced profoundly the limitations imposed by academic culture. I felt an intense pressure to finish on time, I regularly compared myself and my work to others, and I struggled to ask for help despite believing so strongly in the power of collaboration.

This is dominator culture in action, pushing us to achieve through isolation rather than through relationship building. Other master's students finishing their end projects have shared with me similar thoughts. We work largely in isolation, we push ourselves to do more when we need a break, and we constantly adjust our work to fit into an existing structure that doesn't always

fit our strengths, interests, or needs. Academia places greater value on having work that is thorough and valid, rather than on work that is done with love, passion, or joy. We tell students at the start of their process to find a gap in literature to research, rather than to find a gap in their own education to enhance. A mentor of mine told me at the start of my project — no one changes the world with their master's thesis. What matters is how your thesis changes you, because that will impact how you move through the world for the rest of your life. What if this is how we framed all master's projects, all education, or even all research? Instead of evaluating students based on outputs, what if we evaluated them based on how we are able to bring others into our process collaboratively, based on what we learn along the way? To center relationships – to focus on joy, love, and hope in the research process – is a way to create a more inclusive research community.

> "The ability to laugh and enjoy the research process is fundamental to our research practices. If we remain stuck in a space of anger, frustration, or disillusionment, then putting our research efforts toward a greater good does not become possible." (Falcón, 2016, p. 188)

Bibliography

Anthony-Stevens, V., & Matsaw Jr, S. L. (2020). The productive uncertainty of indigenous and decolonizing methodologies in the preparation of interdisciplinary STEM researchers. *Cultural Studies of Science Education*, *15*(2), 595–613. https://doi.org/10.1007/s11422-019-09942-x

Beeman-Cadwallader, N., Quigley, C., & Yazzie-Mintz, T. (2012). Enacting decolonized methodologies: The doing of research in educational communities [Number: 1]. *Qualitative Inquiry*, *18*(1), 3–15. https://doi.org/10.1177/1077800411426414

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new jim code*. Polity.

Blue, G., Rosol, M., & Fast, V. (2019). Justice as parity of participation: Enhancing arnstein's ladder through fraser's justice framework [Number: 3]. *Journal of the American Planning Association*, *85*(3), 363–376. https://doi.org/10.1080/01944363.2019.1619476

Bon, A., Dittoh, F., Lô, G., Pini, M., Bwana, R., WaiShiang, C., Kulathuramaiyer, N., & Baart, A. (2022). Decolonizing technology and society: A perspective from the global south. In H. Werthner, E. Prem, E. A. Lee, & C. Ghezzi (Eds.), *Perspectives on digital humanism* (pp. 61–68). Springer International Publishing. https://doi.org/10.1007/978-3-030-86144-5_9

Boyd, A. (2021). Intersectionality and reflexivity—decolonizing methodologies for the data science process [Number: 12]. *Patterns*, *2*(12), 100386. https://doi.org/10.1016/j.patter.2021.100386

Caxaj, C. S. (2015). Indigenous storytelling and participatory action research: Allies toward decolonization? reflections from the peoples' international health tribunal. *Global Qualitative Nursing Research*, *2*, 233339361558076. https://doi.org/10.1177/2333393615580764

Chan, L., Hall, B., Piron, F., Tandon, R., & Williams, L. (2020, July). *Open science beyond open access: For and with communities. a step towards the decolonization of knowledge*. Ottawa, Canada.

Chin, M., Beckwith, V., Levy, B., Gulati, S., Macam, A. A., Saxena, T., & Suwarningsih, D. P. S. (2022). Navigating researcher positionality in comparative and international education research: Perspectives from emerging researchers [Number: 2]. *International Education Journal: Comparative Perspectives*, *21*(2). https://openjournals.library.sydney.edu.au/IEJ/article/view/15891

Collins, P. H. (2016). Black feminist thought as oppositional knowledge [Number: 3]. *Departures in Critical Qualitative Research*, *5*(3), 133–144. https://doi.org/10.1525/dcqr.2016.5.3.133

Datta, R. (2018). Decolonizing both researcher and research and its effectiveness in indigenous research [Number: 2]. *Research Ethics*, *14*(2), 1–24. https://doi.org/10.1177/1747016117733296

Eseonu, T., & Duggan, J. (2022). Negotiating cultural appropriation while re-imagining coproduction via afrofuturism [Number: 1]. *Qualitative Research Journal*, *22*(1), 96–107. https://doi.org/10.1108/QRJ-06-2021-0060

Evans, M., Miller, A., Hutchinson, P., & Dingwall, C. (2014, June). Decolonizing research practice: Indigenous methodologies, aboriginal methods, and knowledge/knowing. In P. Leavy (Ed.), *Oxford handbook of qualitative research* (pp. 179–191). Oxford University

Press. https://www.researchgate.net/publication/261872363_De-Colonizing_Research_Practice_Indigenous_Methodologies_Aboriginal_Methods_and_KnowledgeKnowing

Falcón, S. M. (2016). Transnational feminism as a paradigm for decolonizing the practice of research: Identifying feminist principles and methodology criteria for US-based scholars [Number: 1]. *Frontiers: A Journal of Women Studies*, *37*(1), 174. https://doi.org/10.5250/fronjwomestud.37.1.0174

Fish, L., Flavell, M., & Cunningham, E. (2022). Reimagining communities of practice: Using relational frameworks to disrupt assumptions and inequity [Number: 2]. *International Education Journal: Comparative Perspectives*, *21*(2). https://openjournals.library.sydney.edu.au/IEJ/article/view/15887

Franklin, R. S., Delmelle, E. C., Andris, C., Cheng, T., Dodge, S., Franklin, J., Heppenstall, A., Kwan, M.-P., Li, W., McLafferty, S., Miller, J. A., Munroe, D. K., Nelson, T., Öner, Ö., Pumain, D., Stewart, K., Tong, D., & Wentz, E. A. (2023). Making space in geographical analysis [Number: 2]. *Geographical Analysis*, *55*(2), 325–341. https://doi.org/10.1111/gean.12325

Gasparotto, M. (2016). Digital colonization and virtual indigeneity: Indigenous knowledge and algorithm bias [Medium: application/pdf Publisher: Rutgers University]. https://doi.org/10.7282/T3XG9TFG

hooks, b. (2003). *Teaching community: A pedagogy of hope*. Routledge.

Jama, H. (2023, March 31). *The way i approach technology cannot be separated from black womanhood* [Null and void]. Retrieved April 19, 2023, from https://nullandvoid.substack.com/p/attention-complexity-and-humility

Kimmerer, R. W. (2013). *Braiding sweetgrass* (First edition). Milkweed Editions.

Kirkness, V. J., & Barnhardt, R. (2001). First nations and higher education: The four r's - respect, relevance, reciprocity, responsibility. In R. Hayhoe & J. Pan (Eds.), *Knowledge across cultures: A contribution to dialogue among civilizations*. Comparative Education Research Centre, University of Hong Kong. https://www.afn.ca/uploads/files/education2/the4rs.pdf

Meital, K., & Jason, M. (2022). *Language and coloniality: Non-dominant languages in the digital landscape*. Pollicy. https://pollicy.org/wp-content/uploads/2022/08/Languages-Coloniality-Report.pdf

Mingus, M. (2023, March 16). *Pods: The building blocks of transformative justice & collective care* [SOILTJP]. Retrieved October 10, 2023, from https://www.soiltjp.org/our-work/resources/pods

Musara, E., Grant, C., & Vorster, J.-A. (2021). Inclusion as social justice: Nancy fraser's theory in the south african context [Series Title: Springer International Handbooks of Education]. In C. A. Mullen (Ed.), *Handbook of social justice interventions in education* (pp. 39–58). Springer International Publishing. https://doi.org/10.1007/978-3-030-35858-7_107

Nathan, L. P., Kaczmarek, M., Castor, M., Cheng, S., & Mann, R. (2017). Good for whom?: Unsettling research practice. *Proceedings of the 8th International Conference on Communities and Technologies*, 290–297. https://doi.org/10.1145/3083671.3083685

Ndlovu-Gatsheni, S. J. (2019). Provisional notes on decolonizing research methodology and undoing its dirty history [Number: 4]. *Journal of Developing Societies*, *35*(4), 481–492. https://doi.org/10.1177/0169796X19880417

Parthasarathy, S., & Stilgoe, J. (2023, March 19). *The politics of expertise and retelling the story of racism in the pulse oximeter ft. amy moran-thomas* (No. 32). https://shobitap.org/the-received-wisdom/2023/3/17/episode-32

Quijano, A. (2000). Coloniality of power and eurocentrism in latin america [Number: 2]. *International Sociology*, *15*(2), 215–232. https://doi.org/10.1177/0268580900015002005

Rosa, H. (2017). (parity of) participation – the missing link between resources and resonance. In B. Bargu & C. Bottici (Eds.), *Feminism, capitalism, and critique* (pp. 157–166). Springer International Publishing. https://doi.org/10.1007/978-3-319-52386-6_9

Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. Zed Books ; University of Otago Press ; Distributed in the USA exclusively by St. Martin's Press.

Streck, D. R. (2021). Transdisciplinarity as a decolonizing research practice: A latin american perspective. *Diálogos Latinoamericanos*, *29*, 88–100. https://doi.org/10.7146/dl.v29i0.120252

Sultana, F. (2018). The false equivalence of academic freedom and free speech: Defending academic integrity in the age of white supremacy, colonial nostalgia, and anti-intellectualism [Number: 2 Section: Interventions]. *ACME: An International Journal for Critical Geographies*, *17*(2), 228–257. Retrieved April 28, 2023, from https://acme-journal.org/index.php/acme/article/view/1715

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data [Number: 6]. *Environment and Planning D: Society and Space*, *34*(6), 990–1006. https://doi.org/10.1177/0263775816633195

Turner Lee, N. (2018). Detecting racial bias in algorithms and machine learning [Number: 3]. *Journal of Information, Communication and Ethics in Society*, *16*(3), 252–260. https://doi.org/10.1108/JICES-06-2018-0056

A | Co-Creation Sessions

The goal of these co-creation sessions was to de-center myself in the construction of a framework based on decolonizing methods. I come with the privilege and bias of 1) growing up in the U.S., 2) speaking English as a first language, and 3) being educated in a prestigious university in the Netherlands. These sessions were a way to expand my own perspective and assumptions by consulting those who have a different global understanding of decolonization.

A.1 Reaching Out to Potential Experts

I considered the people I talked with as though they were experts informing my project, as they are experts of their own experiences. To find people willing to participate, I looked at all of the student associations at TU Delft (as listed on the webpage Student Associations) and sent emails to those associations that bring together students belonging to a particular culture or region of the world that has a history of colonization by the Dutch. I excluded associations that target all international students, are only for European students, are not organized within TU Delft, or that focus on religion (rather than a region or country). For those associations through which I had a personal connection, I reached out to my contact directly, or asked someone I knew to pass along my request. Associations for which I did not know anyone involved, I emailed the contact provided on their association website.

Included:

- ABC Compas, Dutch Caribbean: Aruba, Bonaire and Curaçao
- LATITUD, Latin American – many formerly colonized countries
- TUDASA, African - many formerly colonized countries
- PPI, Indonesian - colonized by Dutch
- SUBEST, Surinamese - colonized by Dutch
- ISA, Indian - colonized by Dutch
- D.I.S.A. Alborz (Iranian) - colonized by Dutch
- ACSSNL (Chinese) - colonized by Dutch
- DTSO (Taiwanese) - colonized by Dutch

After reflecting on my process, I decided to include MSA 'Ibn Firnas' – a religious association – as well. Indonesia has the largest Muslim population in the world and considering that Muslim stories are overwhelmingly inaccurate or otherwise excluded in many spaces in the U.S., I think it can be fair to consider it an underrepresented perspective. I questioned how much of my choice to exclude "religious" associations was because I felt uncomfortable thinking about engaging with Muslim students. Although this may be partly due to my lack of experience or knowledge of the culture around this religion, I suspect it may have come even more from my own internalized bias growing up in a culture that demonizes Muslim peoples. To counteract this bias, I decided to include this association.

Excluded:

- CSA-EUR, an association based out of Erasmus University in Rotterdam
- Tandem Delft, for all internationals
- Delft United, for all internationals
- ESN Delft, targeting Erasmus scholars with a European focus
- ISSTUD, now known as D.I.S.A. Alborz (included)
- AEGEE Delft, for all internationals, has a European focus
- BEST, for European students

A.2 Protocol for Co-creation Sessions

Estimated time: 1 hour
Location: ideally in person, flexible depending on the comfort of the participant
Other context notes: provide some kind of snack, food, or gift as a thank you for their time and efforts. Remind myself and expert that this is a discussion and let the conversation go wherever it goes. Ask whether the person is willing to be recorded, and take notes. Below is guidance for conversation topics, but constantly invite your co-creator(s) to ask questions and share their thoughts. Respond to their input actively.

Introduction

- Who I am: name, master's program, place of origin
- Who are you? Name, potential common ground (about studies, hobbies, place of origin, food, etc)
- Informed consent – review together, answer questions, sign
- If consent is given, start recording

Project Background

- Purpose of this research project: *The overall goal of this project is to understand how to make the voices of people marginalized in academia heard, respected and included within the data science community at TU Delft. Building on momentum in the data science community to decolonize data science, and I will explain a bit how I understand that effort/momentum but am also curious to know more what that means to you.*
- Purpose of this session: *to get your feedback on how I am approaching my research, see if there's things I am missing because I am an "outsider" of both the data science community, and I am not really from an underrepresented perspective in academia (can share about my own positionality)*
- Open invitation to ask questions about my project: *do you have any initial thoughts before I share a bit more, or anything you'd like me to clarify?*

Define: decolonization

My understanding: *in the context of the U.S., decolonization is centered around empowering and uplifting indigenous voices and communities, and about reparations; from literature I've been reading its about centering indigenous perspectives and understanding role of colonialism in shaping societal institutions so that we can recover (or disrupt) them*

- **How do you define decolonization, specifically efforts to decolonize academic spaces or research?**
- **Where at TU Delft do you see the legacy of colonialism most visibly?**
- Alternate phrasing: what are cultural norms or structural/institutional parts of TU Delft that have come in conflict with your background, or led a feeling of exclusion?

Define: marginalization/underrepresented people

Purpose of this question: *I am wondering what language is most appropriate to use to talk about the people who usually are not included in academic discourses*

- Who do you think of/who would you include when I say: underrepresented, minority, marginalized?
- How would you identify, or how would you want to be identified?

Explore: framework concepts and other indicators

Validating ideas in my existing framework: *The project will look at how social and cultural identities and relationships to others impact a person's ability to express their perspective in the community, and whether they feel like their perspective is represented and reflected back to them from the community. I can draw a simplified version of the framework on a piece of paper.*

- Any initial thoughts?
- Are the concepts or connections between concepts that I am incorporating relevant in your experience? Specifically: relationships and the connection to reciprocity, voice and the connection to representation, perspective
- Are there things I am missing?
- How would you determine whether the inclusion of marginalized voices is successful or not?

A.3 Notes from Co-creation Sessions

Initially, as I began reaching out to student organizations from different cultural backgrounds, I was extremely cognizant of the fact that I asked for their help but could offer nothing in return. This felt uncomfortable for me and I delayed sending emails for multiple days because I was second guessing how to phrase my request in a way that was both open and respectful. It was much easier to connect to associations through the people I know directly. But that is the power of networks. Networks should be reciprocal, and I hope that those from whom I asked for support feel as though they can also ask me for things in return in the future.

As I suspected, I did not get much response from the cold emails I sent. Although three organizations did respond and pass my message along to their members, the conversation did not continue after that initial reply. The two students I did end up meeting with were both through direct relationships. They were both interesting and enlightening conversations. To respect their privacy, I don't specify which student association they are from. As a way to say thank you for their time, I also provided both with food.

Below is a summary of the notes I took from both conversations.

Defining decolonization: it's about erasing the gap created by the colonizer and acknowledging the exploitation that happened, including violation of rights. Decolonization must be an ongoing process. As a concept, decolonization is very abstract – how can you make it concrete? What kinds of actions are decolonizing?

Diversity: efforts to make TU Delft more inclusive are hit or miss, often conversations about "diversity" here are about diversity within the Netherlands or Europe – not other ethnic backgrounds or other parts of the world. People here are unsure how to navigate diversity related to "people of color." Experiencing racism makes them hesitant to approach new people, because they don't know how they will be treated in a group. It's hard to enter a world you feel like is against you.

International students and belonging: Non-EU students are not that welcomed. Feels like Dutch and EU students are prioritized more, then maybe students from the US and Canada. There is a lot of emphasis placed on Dutch and EU research, nationalist bias. International students are not as listened to and have less access and opportunities. International students also have less connections; it's harder to form. Feel like you have to adapt when you come here as an international.

Language: language impacts our ability to connect, can also affect our ability to express ideas. There's a double standard with language—there is little effort put into understanding certain accents or languages from other parts of the world, but there is an expectation that everyone will speak English (or Dutch to a certain extent).

Access: accessibility is limited because resources are not well known or advertised. People in gatekeeping roles such as study advisor can limit access to certain resources by not encouraging you to pursue them, not informing you about them (withholding information), or because you need their approval to get those resources. These types of people don't have the capacity to address challenges of race, ethnicity and identity.

Imagining a more inclusive environment: equal opportunities based on what people need. Eradicating discrimination. An environment where you have certainty that everyone feels inequity is wrong. Feeling trust that you'll be supported when you speak up against discrimination. Putting more people of color in gatekeeping roles, representation in staff – but not just for show, having actual diversity in positions of power. A proper diverse environment is one where people know how to engage with other cultures.

B | Participation Interest Form

This appendix includes a copy of the form I sent out to recruit students and a summary of the responses I received. The form was shared via QR code in classrooms I visited and through a link in WhatsApp groups for first and second year computer science bachelor's students. The form was only used to recruit and scope bachelor's and master's students and was not sent to any PhD candidates, postdocs, or professors. Below is a blank form:

Inclusion and Diversity in the Data Science Community: Participation Interest Form

Thank you for your interest in being a part of my master's thesis project. I (Tian Qing Yen) am a student in Communication Design for Innovation at the TU Delft.

The purpose of my project is to understand the extent to which different perspectives and worldviews are represented and included in the data science community at TU Delft. I am looking for participants who:

- Study data science, do research involving data science, or have an interest in becoming a data scientist
- Are in a masters or bachelors program at EEMCS or TPM

Participation will involve a discussion (informal interview) about your relationship to your work and education in data science and your experience expressing your individual perspectives (e.g. values, worldview, experiences, social identities) in this community.

I will only use the information collected in this form to identify participants for this project that have different backgrounds and to communicate about this project. The information you provide will only be accessible to me and my supervisory committee. If you have any questions or concerns, send me an email.

* Required

1. Are you willing to participate in a ~ 1 hour discussion? *

- ○ Yes
- ○ No

2. What is the best way to contact you (e.g. email address, WhatsApp number, etc)? *

3. What study program and track are you currently following? **Please be specific**
Example: 2nd year master's in computer science, data science and technology track

4. Do you identify as a person of color (i.e. non-white)?

- ○ Yes
- ○ No
- ○ I don't know
- ○ Prefer not to say
- ○ Other

5. Do you consider yourself to be a minority at TU Delft?
Being part of a minority means having an identity, background or status that could lead to differential treatment (such as prejudice or discrimination) compared to the dominant groups at TU Delft.

- ○ Yes
- ○ No
- ○ I don't know

6. If *yes*, complete the sentence:
I consider myself a minority because of my ... (select all that apply)

- [ ] Gender
- [ ] Sexuality
- [ ] Race or ethnicity
- [ ] Religion
- [ ] Socioeconomic status
- [ ] Disability
- [ ] Non-EU status
- [ ] Political views
- [ ] Language(s) I speak/don't speak
- [ ] Other

7. Do you have any questions, concerns, comments?

Summary of Interest Form Responses

I prepared this form anticipating that I would receive more responses than I could follow up with; however I only received 4 total. One response was a "no" to participating, and thus they did not fill out the remainder of the form. I contacted and interviewed the other three students who responded. Below is a summary of these three respondents' answers to the scoping questions in the form:

4. Do you identify as a person of color (i.e. non-white)?

More Details

Yes	2
No	1
I don't know	0
Prefer not to say	0
Other	0

5. Do you consider yourself to be a minority at TU Delft?
Being part of a minority means having an identity, background or status that could lead to differential treatment (such as prejudice or discrimination) compared to the dominant groups at TU Delft.

More Details

Yes	3
No	0
I don't know	0

6. If *yes*, complete the sentence:
I consider myself a minority because of my ... (select all that apply)

More Details

Gender	0
Sexuality	1
Race or ethnicity	2
Religion	2
Socioeconomic status	0
Disability	0
Non-EU status	1
Political views	0
Language(s) I speak/don't speak	0
Other	0

C | Interview Protocol

This appendix lays out the interview protocol I used for leading semi-structured discussions with participants. The protocol is written for 45 minutes, however the interviews lasted anywhere between 25 and 75 minutes. The interview was split into three sections. As much as possible, I tried to cover at least some of the questions in each section when time was short.

Section 1 (Appendix C.1) focused on easing into the conversation and getting to know the participant. Before recording, I began with an introduction of myself and my project and went over the consent form to answer any questions the participants might have. I started the interview by inviting participants to reflect on the scope of my project (the data science community) with me. The rest of the questions in this section covered a few main topics. I prioritized asking the questions in bold, following up with other questions if there was time or it fit with the direction the conversation was moving.

Section 2 (Appendix C.2) consisted of a drawing activity and follow-up questions. I provided blank paper and colored pencils for participants to use. The amount of time participants spent drawing varied, though it was usually less than 5 minutes. Some participants explained as they drew, others I asked to explain their drawing after they finished. Depending on what was in the drawing, I asked some follow-up questions to learn more about what they included. The protocol lists some of the questions I asked in follow-up.

The last section (Appendix C.3) was a collaborative brainstorming activity to reflect on resources and processes in the community. I provided participants with a list of resources and processes (Figure C.2). I explained that I was curious to learn more about what resources and processes are in the community and whether people have access to them or know about them. Many participants had a lot to say about the list without much additional prompting. However, the protocol contains some questions I used to encourage comments when participants had less to say. When time allowed, I finished the interviews with a brief reflection on the conversation itself, inspired by the idea that interviews *are* opportunities for participants to reflect.

C.1 Section 1 - Questions

Estimated time: 20 min

Project Introduction

1. Brief introduction of my project:
    - About me (name, program)
    - Goals of project (see informed consent info sheet)
    - Go through informed consent together (and sign if not already signed digitally)
2. Invite participants to ask questions proactively and remind them that this is a discussion
    - Can discuss together how to interpret the questions I ask
    - Feel free to share how questions make you feel or about your experience answering them (provocative, surprising, boring, whatever it is)
3. Begin recording

Scoping the Data Science Community Together

This project is about the "TU Delft Data Science Community," which I am interpreting as anyone affiliated with TU Delft who is studying or doing research involving data science or using data science methods. I'm curious if you have a similar or different definition of the data science community:

1. **How would you define who is part of a data science community at TU Delft?**
2. Are there groups missing from the scope I've provided that you would include?
3. Are there people I've included who you wouldn't consider a data scientist?

Now that we've clarified who is part of the community together, we can focus on those people when thinking about the rest of the conversation.

Participant Background: Agency, Research, and Relationship to Data Science

1. **To confirm, you are a [title/position] in [faculty]?**
2. Do you have any other formal or informal roles?
3. Can you briefly describe your research? OR (For students) do you have a specific track or area of focus?
4. **Why did you choose to study or do research in data science (or in your particular area of research)?**
5. What other factors besides personal interest influenced your choice to do this research/study?

Representation of Values and Perspectives

1. **What values do you have as a data scientist?**
2. **Do you like these values are reflected or shared by others in the community?**
3. (For students) Are these values the ones you've been taught or learned? OR (For professors) What values do you try to teach in your classes?
4. Are these values (that you have or that are shared in the community) explicit or implicit?

Identity and Reflexivity

1. **What aspects of your identity, if any, impact how you approach your research or approach your work/studies?**
    - If participant confused by "identity", provide visual from Figure C.1 with various identity types
2. **What opportunities do you have to share/reflect/talk about the ways your identity shapes the assumptions you make in your research?**

Figure C.1: Drawing to represent various identities. This was a visual I provided to participants to prompt them to consider various ways to interpret "identity" if they seemed stuck.

C.2 Section 2 - Drawing Activity

Estimated time: 15 min

Drawing Position and Relationships

Prompt: *Considering our earlier definition of the data science community, draw how you are connected to this community*

- Activity is meant to be a conversation starter, so can explain as you are drawing and it doesn't need to be perfect in any way
- Can use different shapes and sizes, texture/shading or color to represent different qualities
- (If participant needs more guidance) consider including various individuals you are connected with, different courses, projects, lab groups, faculties. Can also include processes or events

Follow-up Questions: Pick 2-3 depending on time and content of the drawing:

- Why are certain relationships or connections you drew represented differently (with color, texture, line-style, etc.)?
- Where would you consider the center of this community?
- Where would you put yourself in this drawing?
- How might your ideal version of this "community" look different? OR What do you want to change about how you have drawn the community?
- Where do you feel the most support in this community?
- I noticed you didn't include [something] in your drawing. Why is that? Where might that [something] be?

C.3 Section 3 - Reflecting Together

Estimated time: 10 min

Brainstorming Activity: Community Resources, Processes and Transformations

Prompt: *I want to get a sense of what resources are present in the community and who has access to them, as well as where transformations or processes might be happening. I have a list (see Figure C.2) that I've been expanding on from each discussion.*

- Is there anything specific you think is missing from these lists?
- Is there anything on this list that you are unfamiliar or don't know what it is? OR Are you familiar with [specific thing on the list I think might be relevant to them]?
- Do you feel like you can access [resource on the list]? OR Do you feel like you have influence over [process/transformation on the list]?
- What do you think *should* be added to the list? What resources do you wish the data science community had or what changes do you want to see happen in the community?

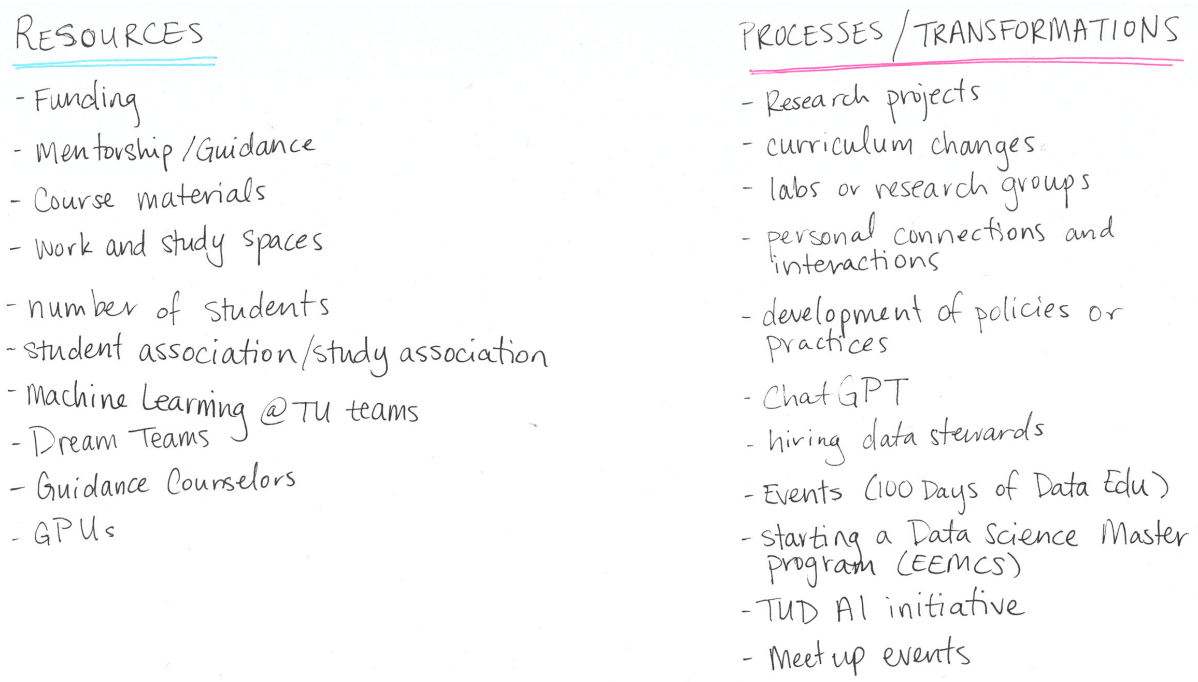


Figure C.2: List of resources and processes and transformations. A physical version of this list was shown at all interviews, except during the one interview that occurred online.

Reflecting on the Conversation

1. What changes do you see happening in the data science community or at TU Delft as a whole?
2. Do you have any questions for me or about my project?
3. Why did you agree to participate in this project?
4. What did you get out of this conversation? Did anything surprise you? Did you learn something?

D | Visuals from the Design Process

Figure D.1: Abstract design ideas. The initial abstract painting I created to come up with ideas for how to address various complexities in the community. Most of the words were ideas taken directly from interviews, though a few are from artwork shared by or conversations with friends.

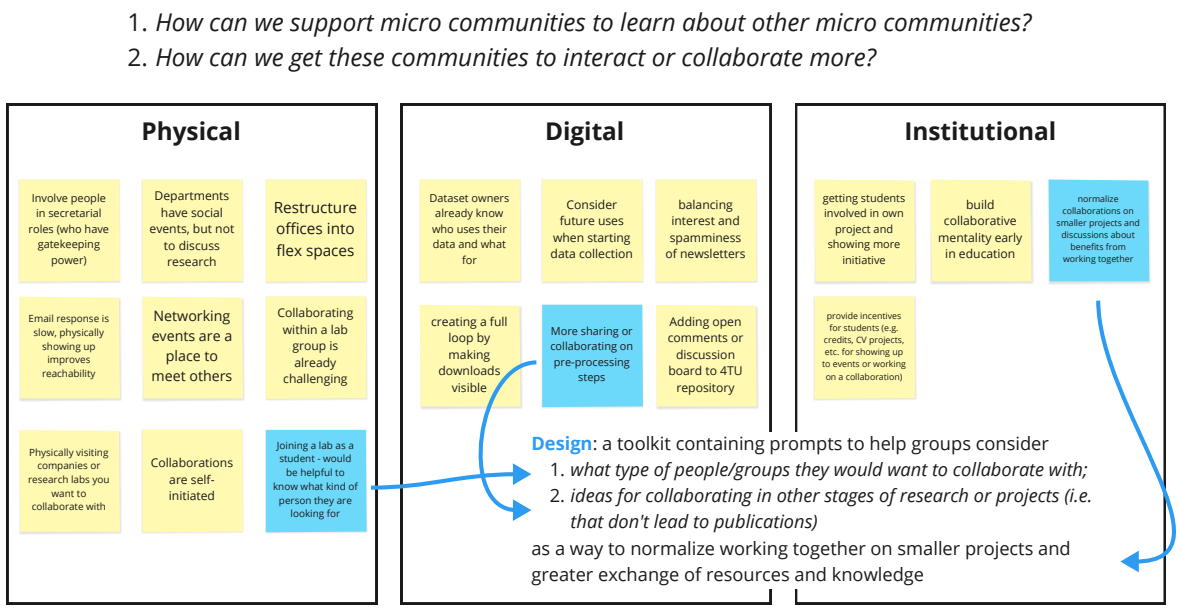


Figure D.2: Ideas from second design brainstorm session. This board shows the ideas that came out of a collaborative brainstorming session. The questions at the top of the board were the prompts I gave participants, and I guided the conversation to consider physical design ideas, digital ideas, and institutional ideas or changes to answer these questions. My key takeaways are highlighted blue and point to the parts of my final design idea that they influenced.

E | Risk and Data Management Plans

This appendix contains materials prepared for the Human Ethics Review Committee. It contains copies of the informed consent used for my co-creation sessions and interviews (described in Section 3.1.2 and Section 3.2 respectively), and the risk mitigation matrix I prepared prior to starting my research. Below is the order in which these documents appear in this appendix.

1. Informed Consent for Co-Creation Sessions
2. Informed Consent for Interviews
3. Risk Mitigation Matrix

Project and Co-creation Session Information

Thank you for your willingness to support this project on **Decolonizing the data science community through meaningful inclusion of underrepresented voices**, being carried out by Tian Qing Yen, a master's student in Communication Design for Innovation at the TU Delft.

The overall goal of this project is to understand how to make the voices of people marginalized in academia heard, respected and included within the data science community at TU Delft. The project will look at how social and cultural identities and relationships to others impact a person's ability to express their perspective in the community, and whether they feel like their perspective is represented and reflected back to them from the community.

The purpose of this co-creation session is to advise and provide feedback on the framework for this project. The framework will be used to shape interview questions and define abstract concepts such as representation, respect, and inclusion. I believe it is important that these concepts are defined in a way that is not just based in theory, but also accurately reflects the reality of people at TU Delft that come from (historically) marginalized backgrounds. This session will last around **1 hour** and is meant to be a discussion. I will ask your perspectives on the framework and invite you to share as much or as little about your experiences as you feel is relevant.

This discussion and your input will **NOT** be analysed and will only be used to inform the background and framework of this project. I will take anonymous notes during the discussion that I will send back to you so that you can review, revise, or redact any (or all) parts of your input. These notes will only be available to myself and my supervisory committee. If you consent, I will also record this session to listen back to as I am revising the framework. This audio would be stored on a secure project drive only available to myself and would be deleted after the completion of the project. I will not ask for any personally identifiable information, except for your name on this consent form. You may also choose if you would like your contributions to be acknowledged, or if you would prefer to remain entirely anonymous.

If you have any questions, concerns or complaints about the research project or this session, please email Tian Qing Yen at [] or send a message to the supervisor of this thesis at [].

Consent for participation in a co-creation session

1. I have read the information above, have been able to ask questions about this session and the project, and my questions have been answered to my satisfaction.

2. I consent voluntarily to participating in this session and understand that my input and responses will NOT be used for analysis, only for revising and validating the framework and background of this project.

3. I understand that notes will be taken about the discussion, that these notes will omit any potentially identifying details to protect my anonymity.

Name:

Date:

Signature:

Project Information for Interviews

You are being invited to participate in a project on **Decolonizing the data science community through meaningful inclusion of underrepresented voices**. This project is being done by Tian Qing Yen, a master's student in Communication Design for Innovation at the TU Delft.

The purpose of your participation in this interview is to understand the extent to which different perspectives and worldviews are represented and included in the data science community at TU Delft. This interview is a **discussion** that will last around 1 hour. Tian Qing (the main researcher) will ask prompting questions about your own perspectives (values, worldview, social identities), your ability and comfort in expressing your perspectives, and your relationship to your education/work in data science and to other data scientists at TU Delft. If answering some questions is difficult or uncomfortable, you can share as much or as little as you would like and can choose to skip questions or end the conversation at any time.

Your responses will be transcribed and pseudonymized by removing all identifying information that could connect it back to you as an individual, and instead connecting the transcript to a random code. Only the main researcher will have access to this code. This will allow her to send the transcript back to you after completion so that you may review, revise, or redact any (or all) parts of your answers. This is also an opportunity to shape how you would like your story and experience to be represented.

With your consent, audio of the interview will be recorded. The audio file will only be listened to by the main researcher and will be deleted immediately after the transcript is made. The pseudonymized transcript will be shared with the research team (Tian Qing and her thesis committee). With your consent, fully anonymous quotes from the transcripts may also be used when describing results in the final thesis. Consent for the latter is not required to participate in the interview.

Your responses will be qualitatively analysed to better understand inclusion in the data science community and to inform the development of a tool or process with the purpose of amplifying a greater diversity of voices in the community. Only fully anonymous results will be published in the master's thesis of Tian Qing and corresponding defence.

Since this project requires collecting personal information, there is a risk of a data storage breach or potential re-identification. To the best of our ability your answers will remain confidential. We will minimize any risks by storing personal identifiable information and sensitive data in a secure project drive managed by TU Delft and only accessible to the main researcher and her supervisory committee.

Your participation in this study is entirely voluntary: you can withdraw at any time and request all data and responses be deleted prior to the completion of the thesis associated with this project (projected for end of September 2023). Your personal information (name and email) collected in this consent form will not be shared or used for any purposes beyond communication about this project and will be destroyed upon completion of the project.

Any questions, concerns or complaints about the research project or processing of data can be directed to Tian Qing Yen by email to [] or to the supervisor of this thesis, by sending an email to [].

III. Risk Assessment and Mitigation Plan

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
A: Partners and collaboration						
1. Will the research be carried out in collaboration with additional organisational partners such as: • One or more collaborating research and/or commercial organisations • Either a research, or a work experience internship provider[1] [1] *If yes, please include the graduation agreement in this application*		**X**				
2. Is this research dependent on a Data Transfer or Processing Agreement with a collaborating partner or third party supplier? *If yes please provide a copy of the signed DTA/DPA*		**X**				
3. Has this research been approved by another (external) research ethics committee (e.g.: HREC and/or MREC/METC)? *If yes, please provide a copy of the approval (if possible) and summarise any key points in your Risk Management section below*		**X**				
B: Location						
4. Will the research take place in a country or countries, other than the Netherlands, within the EU?		**X**				
5. Will the research take place in a country or countries outside the EU?		**X**				
6. Will the research take place in a place/region or of higher risk – including known dangerous locations (in any country) or locations with non-democratic regimes?		**X**				
C: Participants						
7. Will the study involve participants who **may** be vulnerable and possibly (legally) unable to give informed consent? (e.g., children below the legal age for giving consent, people with learning difficulties, people living in care or nursing homes,).		**X**	All participants will be affiliated with the TU Delft and able to give informed consent. Since some bachelor's students enrolled may be under the age of 18, these students will be excluded from the research.	I will confirm prior to scheduling an interview and obtaining informed consent that the student is over the age of 18.		
8. Will the study involve participants who **may** be vulnerable under specific circumstances and in specific contexts, such as victims and witnesses of violence, including domestic violence; sex workers; members of minority groups, refugees, irregular migrants or dissidents?	**X**		Since I am asking about social identity and inclusion, there may be members of a minority group who participate. However, participants of this research will have the opportunity to self-identify their	I will remind participants in the informed consent prior to their participation, during their participation, and after completion of the transcript that they may choose not to share about their social identity, or to		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
			minority or vulnerable group membership and do not need to share this information. There is still a risk that participants in this research are targeted due to an implied expectation that participants will be from minority backgrounds.	remove that data from analysis or results. 1-on-1 interviews and flexibility for the participant to choose where and when we meet will reduce the chances of them being identified as a participant of the study. Participant data will be stored only in a secure project drive and pseudonymized to limit the risk of re-identification after the interview. See more details about data storage in the DMP and in question 10.		
9. Are the participants, outside the context of the research, in a dependent or subordinate position to the investigator (such as own children, own students or employees of either TU Delft and/or a collaborating partner organisation)? *It is essential that you safeguard against possible adverse consequences of this situation (such as allowing a student's failure to participate to your satisfaction to affect your evaluation of their coursework).*		**X**	Using recruitment methods such as going through a classroom (relying on professors to encourage students to participate) may pressure students into joining.	I will emphasize that participation is voluntary when I explain the project and will also (when possible) go to the classroom myself so that the message is not coming from a professor.		
10. Is there a high possibility of re-identification for your participants? (e.g., do they have a very specialist job of which there are only a small number in a given country, are they members of a small community, or employees from a partner company collaborating in the research? Or are they one of only a handful of (expert) participants in the study?	**X**		Re-identification is possible, since all participants will be members of a community at TU Delft and are likely to know each other (especially at the professor or postdoc/PhD levels). Other identifying factors such as position, faculty or research group and various social identities may also be used to identify an individual in this community. Re-identification carries a risk to an individuals reputation or relationships or could make them a target if they share about a particular social identity or perspective.	Sensitive personal data and personal identifying information will be stored on a secure TU Delft project drive with limited access. Transcripts will be pseudonymized to limit re-identification and all shared or published results or data will be fully anonymous (see DMP for more details). Where possible for analysis, identifying factors will be aggregated so as to protect the anonymity of individual participants. I will also take extra care to be careful in how I describe relationships between people in the small community so as not to reveal identities.	DMP I.3, V.26	
D: Recruiting Participants						
11. Will your participants be recruited through your own, professional, channels such as conference attendance lists, or through specific network/s such as self-help groups		**X**				
12. Will the participants be recruited or accessed in the longer term by a (legal or customary) gatekeeper? (e.g., an adult professional working with children; a community leader or family member who has this customary role – within or outside the EU; the data producer of a long-term cohort study)	**X**		Recruitment will happen through direct contacts I have with professors and their associated networks. For instance, I will rely on professors as gatekeepers to reach students in their course/class/research group. This carries a risk of participants being identified as part of this study due to how they were	Although I will ask participants if they have contacts of others who might be interested in joining this research, I will not share any information about who actually agrees to participate. I will be sure to acknowledge potential biases that may occur from this recruitment method in my report and in relevant metadata for storing or sharing (see DMP for more details). I will be	DMP V.26, V.28	

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
			recruited. It also carries a risk of biasing who participates in the study.	clear with the individuals who I am reaching out to for contacts about my intentions around who I would like to participate.		
13. Will you be recruiting your participants through a crowd-sourcing service and/or involve a third party data-gathering service, such as a survey platform?		**X**				
14. Will you be offering any financial, or other, remuneration to participants, and might this induce or bias participation?		**X**	I will provide some snacks (cookies, chocolate, etc.) as a thank you for participants joining the study. I do not anticipate that this will bias their responses or participation in this research.	Although I will mention bringing a snack, I will not emphasize this point in my recruitment strategy. See also question 13 for more details.		
E: Subject Matter *Research related to medical questions/health may require special attention. See also the website of the CCMO before contacting the HREC.*						
15. Will your research involve any of the following: • Medical research and/or clinical trials • Invasive sampling and/or medical imaging • Medical and *In Vitro Diagnostic Medical* Devices Research		**X**				
16. Will drugs, placebos, or other substances (e.g., drinks, foods, food or drink constituents, dietary supplements) be administered to the study participants? *If yes see here to determine whether medical ethical approval is required*		**X**	As part of building a reciprocal relationship with participants and to appreciate them, I would like to provide snacks (cookies, chocolate, etc.). This food is not part of the research design.	Before opening a food package or placing it in a room, I will ask if the participant has food allergies. I will clearly communicate all ingredients in the food provided, and that consuming the food will be entirely voluntary.		
17. Will blood or tissue samples be obtained from participants? *If yes see here to determine whether medical ethical approval is required*		**X**				
18. Does the study risk causing psychological stress or anxiety beyond that normally encountered by the participants in their life outside research?		**X**	Discussing topics around social identity, inclusion, and (potential) exclusion or oppression may be upsetting for some participants and cause anxiety. Although this is not expected, it is a risk, especially if these are topics the participant does not regularly think about or discuss.	Before and during interview conversations, I will remind participants that they can skip questions or end the interview at any time. I will pay attention to body language to gauge the comfort of the participant and pause or stop an interview if I suspect they are feeling excessive stress or anxiety.		
19. Will the study involve discussion of personal sensitive data which could put participants at increased legal, financial, reputational, security or other risk? (e.g., financial data, location data, data relating to children or other vulnerable groups) *Definitions of sensitive personal data, and special cases are provided on the TUD Privacy Team website.*		**X**	Questions about worldview, perspective, and social identity also may involve discussion of sensitive data such as race, religion, social class, etc. This can pose risks such as reidentification or being treated differently by peers because of a social group status, including risk to reputation or risk of violence, harm or oppression.	For reidentification, see measures discussed in 10.		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
20. Will the study involve disclosing commercially or professionally sensitive, or confidential information? (e.g., relating to decision-making processes or business strategies which might, for example, be of interest to competitors)		**X**				
21. Has your study been identified by the TU Delft Privacy Team as requiring a Data Processing Impact Assessment (DPIA)? *If yes please attach the advice/ approval from the Privacy Team to this application*		**X**				
22. Does your research investigate causes or areas of conflict? *If yes please confirm that your fieldwork has been discussed with the appropriate safety/security advisors and approved by your Department/Faculty.*		**X**				
23. Does your research involve observing illegal activities or data processed or provided by authorities responsible for preventing, investigating, detecting or prosecuting criminal offences *If so please confirm that your work has been discussed with the appropriate legal advisors and approved by your Department/Faculty.*		**X**				
F: Research Methods						
24. Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g., covert observation of people in non-public places).		**X**	Participants will be fully aware of the purpose and design of the study when volunteering.			
25. Will the study involve actively deceiving the participants? (For example, will participants be deliberately falsely informed, will information be withheld from them or will they be misled in such a way that they are likely to object or show unease when debriefed about the study).		**X**				
26. Is pain or more than mild discomfort likely to result from the study? And/or could your research activity cause an accident involving (non-) participants?		**X**				
27. Will the experiment involve the use of devices that are not 'CE' certified? *Only, if 'yes': continue with the following questions:*		**X**				
• Was the device built in-house?						
• Was it inspected by a safety expert at TU Delft? *If yes, please provide a signed device report*						
• If it was not built in-house and not CE-certified, was it inspected by some other, qualified authority in safety and approved? *If yes, please provide records of the inspection*						
28. Will your research involve face-to-face encounters with your participants and if so how will you assess and address Covid considerations?	**X**		Holding face-to-face interviews and focus sessions may bring a risk of Covid infection. This could also	I will follow the Covid-regulations stated by the Dutch government or the university (TU Delft) for all in-		

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
			risk spread of other sicknesses such as the flu or general colds.	person meetings with participants. I will confirm with participants before an interview or discussion, if they are comfortable meeting in person and if they would like additional measures beyond those stated in the regulations to be taken (self-tests, wearing masks, etc.) I will also provide the option to postpone, cancel or hold an interview online if that is more appropriate for changing Covid situations or the needs of the participant.		
29. Will your research involve **either:** a) "big data", combined datasets, new data-gathering or new data-merging techniques which might lead to re-identification of your participants **and/or** b) artificial intelligence or algorithm training where, for example biased datasets could lead to biased outcomes?		**X**	My datasets will be relatively small and entirely qualitative, so no automated merging or aggregation methods will be used.			
G: Data Processing and Privacy						
30. Will the research involve collecting, processing and/or storing any directly identifiable PII (Personally Identifiable Information) including name or email address that will be used for administrative purposes only? (eg: obtaining Informed Consent or disbursing remuneration)	**X**		PII from the Informed Consent forms will be stored and used for purposes of communication. A data breach is a potential risk.	Physical informed consent forms will be digitalized and then the physical versions destroyed. The digital informed consent forms will be stored on a Protected U-drive linked to my NetID that only myself and the supervisory team will have access to.	DMP I.3	
31. Will the research involve collecting, processing and/or storing any directly or indirectly identifiable PIRD (Personally Identifiable Research Data) including videos, pictures, IP address, gender, age etc and **what other Personal Research Data** (including personal or professional views) will you be collecting?	**X**		Interviews and sessions will be audio recorded with Microsoft Teams, with consent from participants. Participants will be asked to describe or share about their professional role within the university and about their individual identity, which may include cultural and educational background, religion, race or ethnicity, gender, or worldview. Questions about identity will be open-ended and participants can choose to share as much or as little as they would like about their identity. Collecting information about identity, alongside other PIRD that a participant may share, carries a risk of re-identification of participants. This could have	Audio files will be stored on a secure project drive only and then destroyed following the creation of a pseudonymized transcript. I will send this transcript back to the participant so that they may also revise or redact any parts of their answers that may contain PIRD they would prefer not to be shared, stored or analyzed. See question 10 for more details about mitigating re-identification.	DMP I.3, IV.10	

			If YES please complete the Risk Assessment and Mitigation Plan columns below.		*Please provide the relevant reference #*	
ISSUE	**Yes**	**No**	**RISK ASSESSMENT – what risks could arise?** *Please ensure that you list ALL of the actual risks that could potentially arise – do not simply state whether you consider any such risks are important!*	**MITIGATION PLAN – what mitigating steps will you take?** *Please ensure that you summarise what actual mitigation measures you will take for each potential risk identified – do not simply state that you will e.g. comply with regulations.*	**DMP**	**ICF**
			an effect on a participant's reputation or relationships if it ends up in the wrong hands.			
32. Will this research involve collecting data from the internet, social media and/or publicly available datasets which have been originally contributed by human participants		**X**				
33. Will your research findings be published in one or more forms in the public domain, as e.g., Masters thesis, journal publication, conference presentation or wider public dissemination?	**X**		Research findings will be published in a Master's thesis which will be made available in a public thesis repository and shared in the final thesis defence. Research findings may be used in a publication.	Only fully anonymized and/or aggregated data will be shared in the results of any form of publication. Consent to use their data for this purpose is asked of participants in the Informed Consent form. Explicit consent to quote directly is also asked.		
34. Will your research data be archived for re-use and/or teaching in an open, private or semi-open archive?	**X**		Some data will be openly archived for re-use and transparency. This could lead to data being used for harmful or unwanted purposes	Only fully anonymized and/or aggregated data will be made open. Participants are asked in the consent form if they would like their data to be destroyed or are okay with its open archival or archival on a project drive with limited access. There will be thorough metadata archived with the research data to prevent misuse due to lack of context, and data will be published under a non-derivative liscence.	DMP V.26. V.29	

F | Design Info Sheet and Activity

Reimagining Collaborative Community

Ideas for building inclusive and supportive relationships in data science

Purpose

This activity is a way to visualize how inclusive your community is of various disciplinary perspectives in data science. It involves identifying the connections that your group already has and reflecting on where you might benefit from strengthening existing relationships or forming new ones.

Who is this for?

Labs, centers, research groups, or interest groups who are involved with data science research. Anyone in your group can facilitate this activity, and it works best if you can get everyone involved!

How to Use

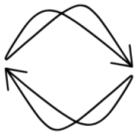
Time: estimated 15-30 minutes

Resources: (in person) a gathering space, a whiteboard or large piece of paper, and pens; (online) an online workspace participants can interact with (e.g., digital whiteboard)

Facilitation: if you are interested in this activity, you can facilitate it! Follow the steps on the next page.

With a small group (2-10), everyone can participate directly in the same conversation(s). With a larger group, the activity may need to be adapted so everyone can contribute.

Why focus on relationships?



Being well connected in academia increases access to resources, knowledge and opportunities

Strengthening relationships can lead to a more connected community

Relationships feel supportive when we have shared experiences and others understand our research

Collaborations are an opportunity to share understanding and experiences

What might it look like to center supportive relationships as the motivation for collaborating in both small and big ways?

Creative Collaborating

What are some benefits of collaborating on projects that are not tied directly to research outcomes?

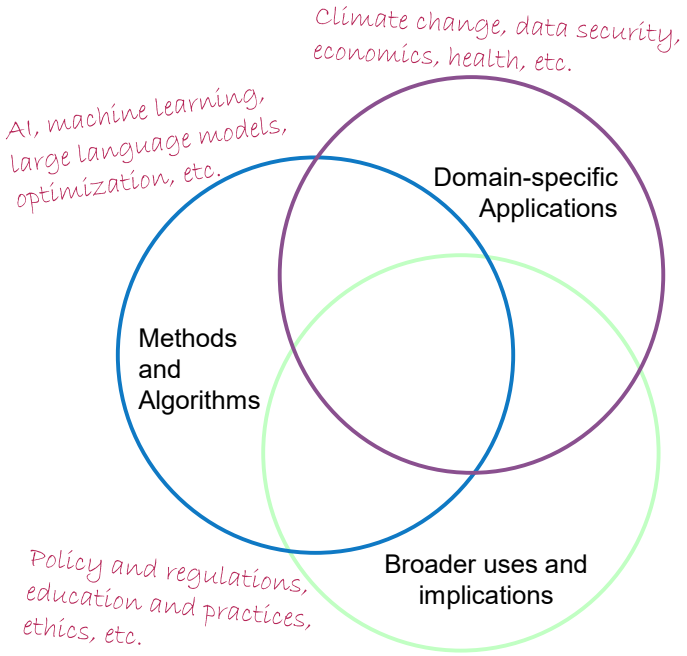
- Strengthen relationships that can lead to future collaborations
- Share resources or processes that can save time in the long run (e.g., methods for cleaning a particular type of data)
- Give or receive support during challenging times
- Provide richer cross-disciplinary experiences for students
- Contribute to the research community or society in different ways (e.g., publishing information for a public audience)
- Increase group visibility

Why is this activity useful?

- Build support within your own group. Reflecting together invites you to learn about how others understand their own work or expertise.
- Align goals and reflect on values in the group. Part of the activity involves discussing what collaborations or connections you might want to strengthen as a group. This creates space to talk about group values and underlying motivations.
- Become more connected and resourced. You might not know what everyone knows. Sharing about existing relationships and resources can increase accessibility for all members of the group.
- Provide ideas for future collaborations within your group or the greater community

Step 1: Map your group together

Recreate the diagram below on a whiteboard, large piece of paper, or shared screen.



Have each person draw or place a mark where they would position themselves within the landscape of data science. *In which circle(s) is your research and/or expertise?*

Briefly share why you put yourself where you did. *What parts of your research (or your skills, knowledge, expertise, etc.) connect to that circle?*

Step 2: Identify growth areas

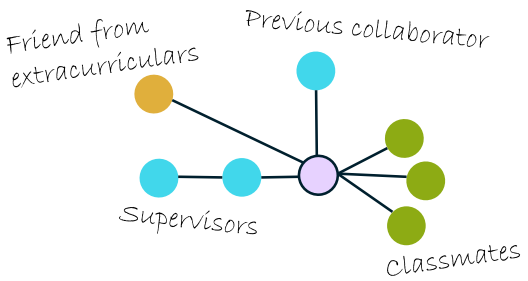
Identify together if there are certain areas that are not represented in your diagram by members of your group.

Discuss where your group might benefit most from having more connections. Look at **Creative Collaborating** for some ideas of different types of "benefits" your group could prioritize.

Step 3: Start from existing connections

Who do you already know that might fill these growth areas? Have everyone think of their existing connections with other people or groups that might fit in the identified growth areas.

Consider relationships that are well-established (i.e., people you could call on if you needed support). For instance, supervisors, existing collaborators, peers with whom you have discussed research challenges or questions, support or administrative roles, others in this group, etc.



Is there someone who is unable to identify any supportive relationships? How might the people in this group become part of their existing network?

Step 4: Ideas to strengthen relationships

Below are a few ideas for how you might strengthen relationships within your group, or between your group and others in the community.

- Ask if someone who has helped you before would be willing to lend similar support to other members of your group
- Invite someone with similar interests to attend an event together
- Proactively share sources, information, expertise, etc. with those who might benefit from it
- Collaborate on projects focused on pre-processing steps (e.g., data management systems, efficiently cleaning data sets, etc.)
- Discuss your values in research with students, supervisors, or existing collaborators
- Share (formally or informally, internally or publicly) about setbacks, challenges, and things that don't work in your research process
- Involve others (especially young professionals) in planning or decision-making processes
- *Don't know anyone who fits the expertise you are missing?* Make a job posting on your website highlighting the kind of student or collaborator you want to work with!