19. Language of algorithms: agency, metaphors, and deliberations in AI discourses

Kaisla Kajava and Nitin Sawhney

INTRODUCTION: WHY SHOULD LANGUAGE MATTER IN AI?

Language is essential to how we as a society make sense of artificial intelligence (AI), while AI technologies increasingly rely on large multilingual language models to make sense of us. Alan Turing's famous quote, "A computer would deserve to be called intelligent if it could deceive a human into believing that it was human" (Turing, 1950), became the basis of his *imitation game* to test a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. This three-actor game, later called the "Turing Test", proposed that if a human evaluator could not reliably judge if another human or machine generated a text-based natural language conversation, the machine would pass the test. While the Turing Test has been both widely influential and criticized over the years, it has captured popular imagination about how we talk about machines and artificial intelligence today. Turing created his test in response to a hypothetical question "Can machines think?" leading him to probe the very ambiguity of defining the meaning of the terms "machine" and "think".

In a later neglected paragraph of the same paper, Turing states:

> The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless, I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

In a sense, Turing was not necessarily predicting that machines would be able to think (like humans) in the future but the idea that thinking machines may one day become uncontroversial. "Our use of the term [thinking] has indeed loosened to the point that attributing thought to even the most basic of machines has become common parlance" (Sandis & Harper, 2017). We may say that "the computer thinks I made a mistake" when it offers typographical or grammatical corrections. Such constructs in everyday language metaphorically associate thinking with machines while implicitly attributing agency to them, rather than the programmers that embed the algorithms and data used for their operation and performance. Searle (1992) makes the distinction between "behavior" and "action", where the former is observable, and the latter performed by an agent with will. In this view, in an AI ecosystem, referring to machines as "performing actions" would falsely transfer a notion of agency to them.

As large-scale multilingual languages models like BERT, GPT4, and LaMDA, and open-source initiatives like BLOOM, become more powerful and prevalent in the design of conversational AI systems, humans are more easily susceptible to anthropomorphizing their capabilities by attributing notions of intelligence and even sentience to them, harkening Turing's original predictions from 1950. The vigorous debates spurred by the claims of a

former Google engineer, Blake Lemoine, about the emergence of sentience in AI were based on his transcripts of an "interview" conducted with LaMDA (Lemoine, 2022). While these claims were widely disputed by most prominent AI researchers and by Google itself (Sparkes, 2022), the popularized debates surrounding such controversies are amplified by public imagination, fears, and aspirations around notions of AI. These notions are shaped in part by the linguistic constructs, terminologies (or their very ambiguity), and wide-ranging discourses about AI circulating through fictional narratives, journalistic reporting, scientific papers, industry reports, and policy statements, among others, in popular culture and educational and professional contexts.

The complexity of AI-based technologies and their increasing prevalence in society makes the challenges of language, terminology, and public discourse especially critical, not just among experts but diverse publics and stakeholders engaged in creating, using, regulating, or deliberating about AI systems. We take an expansive view of "AI systems" as constituted of people, practices, data, technologies, enabling infrastructures, and policies that facilitate the way they operate in the world. Much of the discourse in AI focuses far more on "machine intelligence" and technological aspects of AI systems, rather than the wider ecosystem of human and nonhuman actors, policies, and practices that shape their functioning. Linguistic elements in the form of terminology, metaphors, personification, and so on influence the cultural discourse and political imagination of AI today. As Wyatt (2021) notes, metaphors function more flexibly and dynamically than sociotechnical imaginaries, and they are available to all.

While domain-specific terminology is common in many specialized fields, problematic terms and pragmatic "weak" metaphors (Rehak, 2021) around AI can amplify misunderstandings and expectations while diffusing into other domains (such as politics and public policy) as myths about the unlimited potential, unsubstantiated risks, or autonomy/agency of AI. AI experts and practitioners can both perpetuate and correct these notions, but it also requires critical societal deliberation else the "technical jargon, metaphors, analogy, and linguistic shorthand that may be meaningful to this small group risks being utterly opaque, or worse, misleading" outside this demographic circle, while "rich sources of potentially clearer, more accurate, or more widely accessible metaphors and terminology remain unplumbed" (Bones et al., 2021). Hence, democratizing the discourses around AI requires a critical understanding of its current usage and implications, as well as engaging not just AI experts and practitioners but also wider multidisciplinary perspectives from philosophy, linguistics, social sciences, law, public policy, and other academic disciplines, as well as diverse domains of knowledge in civil society. A multidisciplinary perspective in AI facilitates a more pluralistic, ethical, and inclusive engagement for better informed and responsible outcomes as discussed by Floridi et al. (2018).

The public sector is increasingly embracing algorithmic decision-making and data-centric infrastructures to improve planning and operational efficiency in cities while offering innovative digital services to citizens in areas such as education, healthcare, and urban mobility (Floridi, 2020; Haataja et al., 2020). As such public AI services become more prevalent and affect citizens' lived experiences, we must critically understand their social, political, and ethical implications to examine the rights, risks, and responsibilities for both the providers and recipients of such services, particularly the most vulnerable in society (Crawford, 2021; Sawhney, 2022). Some AI-based systems are being used by governments for biometric surveillance, criminal justice, and other forms of citizen monitoring, posing higher risks

for abuse and unfair incrimination (AI Now, 2018) if they are not made easily transparent, accountable, or their legitimate use challenged by civil society. In a deliberative democracy, the rapid growth of emerging AI technologies and services necessitates new forms of digital citizenship and algorithmic literacy, i.e., having an informed ability to critically examine, evaluate, propose, or contest digital services (Long & Magerko, 2020; Hermann, 2021).

Discourses about AI systems and other algorithmic tools often revolve around issues of trustworthiness (Drobotowicz et al., 2021), transparency (Ananny & Crawford, 2018), or even a kind of mystique (Elish & boyd, 2018), feeding into the conception of AI-based systems as challenging to critically understand. Lack of informed understanding may leave citizens either unwilling to participate or eager to do so unquestioningly, while policymakers and civil servants must contend with the indeterminacy and unexpected outcomes of the AI services they are tasked to provide and oversee. Expanding algorithmic literacy can facilitate a more inclusive, critical, and participatory engagement in AI discourses among civil society.

Our research seeks to understand the positions, narratives, and values embedded in and constructed through ongoing discourses around public sector AI services and policy regulations.[1] In this chapter, we examine critical aspects of how language is used in both AI reports and in multi-stakeholder deliberations on the regulation of AI systems across private and public sectors in the EU. We understand "AI discourses" as discursive conventions of conceptualizing and framing the nature and use of AI technologies. The term "agency" is used in two contexts in this chapter. First, we refer to agency in the linguistic sense of the word "AI" occurring in the semantic role of an agent or the syntactic function of a subject, and how this use of language attributes agency to AI through personification. Second, we talk about agency in a social sense, referring to the capacities, abilities, and opportunities diverse (human) actors have in the deliberations around AI development, policy, and regulation. In this chapter, we examine how both types of agency are embedded in the thematic concerns, metaphors, and personification prevalent in the European Commission's AI Watch reports, and diverse stakeholder deliberations around the proposed EU AI Act. Our focus here is limited to critically examining the linguistic and discourse aspects of these AI reports and stakeholder feedback, while the complex policy implications are not within the scope of our current analysis in this chapter.

MAKING SENSE OF LANGUAGE AND DISCOURSES IN AI

Helen Bones et al. (2021), in their paper "In the Frame: The Language of AI", examine discourses in AI through the lens of historical use, gender, and language to explore how "values, language, and norms are produced – and to contest reductive approaches to AI". Citing the work of Alison Adam (1995), they suggest that the traditional epistemological assumptions around AI ignore power relations and cultural conditions in society while focusing too narrowly on modeling intelligent human behavior. Building on feminist epistemology, Adam asks how AI systems represent knowledge, what kind of knowledge, and whose knowledge (Adam, 1995), recognizing a plurality of views from the tacit, situated, and contextual knowledge embedded among diverse publics. Hence, Bones et al. (2021) elaborate that terminology

[1] This work is part of the *Civic Agency in AI* (CAAI) research project at Aalto University, supported by the Kone Foundation and the Research Council of Finland.

matters, from the "selection of word stems, source languages, and the figures of speech" that encode certain sociocultural and political positionalities in how we describe AI constructs and technologies. In their paper, they examine the historical and cultural dimensions of popular AI terminologies including "data", "memory", "intelligence", and "learning" as case studies to demonstrate the problematic ways in which AI is framed and can be constructively reframed (Bones et al., 2021). The notion of data existing independent of human creation can be reframed as "capta" (Drucker, 2011) or alternatively as "situated data" (Lavin, 2021), embodying active social construction and contextual interpretation of phenomena, rather than a fixed construct inherent in the world and outside of human intervention. Reframing language affects notions such as ownership, exchange, commoditization, and sovereignty of data in society. Using the term "memory", rather than say "virtual storage", can attribute human qualities like reminiscence or agency to machines and conflate it with the structure and neurological functioning of the human brain. Retrieving accurate copies of stored bits is not how human memory operates; it is instead a highly complex process "comparable to living through imagined events again and by that even changing what is being remembered" (Rehak, 2021).

Marvin Minsky in his 1974 article "A Framework for Representing Knowledge" proposed the concept of "frames" as data structures for knowledge representation and reasoning in the early days of AI research (Minsky, 1975); while not analogous to memory, the notion of frames, derived from experimental psychology, contributed to the development of specialized programming languages and expert systems in the 1980s, and more recently, the Semantic Web. Minsky was an early proponent of AI being able to eventually model human brains, thought processes, and intelligence, whether through the notion of frames for symbolic logic and/or "perceptrons" for connectionist networked learning. However, MIT philosopher Hubert Dreyfus (1972) vigorously argued that intelligent human behavior required "common-sense knowledge" that is partly tacit and acquired through embodiment, cognitive development, and cultural practice, which were unrealistic for computers to acquire (Dreyfus & Dreyfus, 1986; Fjelland, 2020). These arguments remain at the heart of debates around the distinction between artificial general intelligence (AGI) and artificial narrow intelligence (ANI) today.

Related to intelligence, a commonly used term in AI is "learning", which also has broader cognitive and pedagogical meaning (Bones et al., 2021). Alexander et al. (2009) examine the many dimensions of learning being tacit, intentional, disadvantageous, interactional, and so on, essentially as a more holistic activity not confined only to rational thinking but also to survival, sociocultural contexts, and engaging how human senses interact with the world. The processes and methods associated with AI-based approaches like statistical machine learning, deep neural networks, or reinforcement learning are undertaken with a far narrower frame of objectives and constraints. Hence, using the metaphor of *learning* uncritically to describe such methods can lead to unrealistic expectations and misunderstanding of the scope of what AI technologies can learn about the world, unlike complex human forms of social, embodied, and continuous situated learning (or unlearning) over time.

Language use hence reveals implicit information about people's perceptions of AI, taking part in constructing the framework of narratives within which AI technologies are used, developed, and regulated. Discourses around AI systems can even be seen as ethical concerns when they exaggerate the autonomy of technology, attributing human-like characteristics to it, and in doing so, creating a "sociotechnical blindness" (Johnson & Verdicchio, 2017). The role of language is crucial in the translation of information between domains of knowledge, both in shaping discourses and in creating transparent and accessible technical documentation.

Language can facilitate literacy between domains and make it accessible or it can obscure understanding if the terminology is either too specialized or too vague, as in the case of AI.

Once deployed and in use, AI systems are not simply machines routinely performing predefined tasks, even when they require no technical maintenance. Instead, they are sociotechnical systems continuously updated in people's perceptions, and the discourse around them is in constant fluctuation. The pace of AI development creates challenges for traditional ways of policymaking, which may often be a long-term strategic process (Dwivedi et al., 2021). Because of this rapid development, stakeholders, including developers and technology companies, are reactively conceptualizing emerging technologies, and in doing so, shaping the language used to refer to those systems. This communicative exchange contributes to the development of both the technology as well as its use, governance, and regulation. Analogously, the AI Act has seen a lot of debate around its definition of AI: the different technologies bundled under the term "AI" may have different impacts and risks, many of them having little in common, while regulators, policymakers, and industry actors struggle to create a mutually agreed-upon and effective regulatory framework.

We examined two corpora on AI Watch reports and stakeholder deliberations in the context of the European Union to study different aspects of the discourse. The corpora, which consist of reports published by the European Commission's AI Watch and multi-stakeholder responses to the proposed AI Act, offer us a lens to unpack the use of metaphors and personification of AI technologies as well as some of the ongoing deliberations around AI regulation in Europe. While the curation and compilation of corpora entail limitations regarding the generalizability of the discursive practices to other contexts, these datasets are valuable for both qualitative and quantitative analysis within their scope of relevance. In our ongoing research, we combine methods from natural language processing (NLP) and critical discourse studies (CDS) to capture information at different granularities. NLP techniques serve as a method for hypothesis generation through search: combined with qualitative methods, computational text processing allows us to discover statistical patterns such as n-grams, named entities and keywords, parse syntactic dependencies to analyze agentic relations between concepts, and summarize long documents to aid in sample selection for further analysis. This information extraction enables us to create and revise working hypotheses for more in-depth qualitative inquiry.

METAPHORS AND LINGUISTIC AGENCY IN AI REPORTS AND POLICIES

Metaphors are prominent linguistic devices used to make sense of the reality around us (Kövecses, 2020). The ubiquity and everydayness of metaphors make them a particularly interesting linguistic element to study, offering a lens to implicit conceptualizations and bringing attention to fluctuations and conventions of discourse, which may often go unnoticed. AI discourse is full of metaphors, which hyperbolize its autonomy and agency (Bones et al., 2020; Rehak, 2021) or proliferate a sense of mythical AI as a benevolent force (Ossewaarde & Gulenc, 2020).

Conceptual metaphor theory (CMT) (Lakoff & Johnson, 2003; Kövecses, 2020) posits that metaphors involve the analogous mapping of one concept to another with the purpose of making sense of complex phenomena and social reality. Thus, a metaphor is not just a

linguistic phenomenon but a cognitive sense-making device. According to the theory, a metaphor involves two conceptual domains: a target domain and a source domain. The target domain is the concept that requires sense-making, while the source domain is the concept used to make sense of the target domain, for example, in the classic example ARGUMENT IS WAR, *argument* is the target domain, and *war* is the source domain, which ties the concept of *argument* to something more tangible. Analogously, "shooting down" or "winning" an argument are realizations of the conceptual metaphor (CM) ARGUMENT IS WAR. Coupled with critical metaphor analysis (CMA) (Charteris-Black, 2004), CMT offers us a critical view of the use of metaphors in society. CMA has its roots in critical discourse studies and aims to uncover covert uses of metaphor. Metaphors may be intentional or purposeful rhetorical devices, or unconsciously motivated (Charteris-Black, 2004), of which the latter type is of interest to us here. Metaphors in EU policy documents have been critically studied in relation to the market economy and economic agents (Dewandre & Gulyás, 2018), but to our knowledge, the metaphors and language of AI have not been explored in this context using frameworks from CMT and CMA. As Dewandre and Gulyás (2018) note in their analysis of metaphor in EU documents, policy reports and other policy-related texts may indeed encode a conscious rhetorical voice, but our main interest in using CMA here is to bring attention to the conceptions and conventions of language underlying AI-related policy.

Beyond metaphors, we can operationalize the concept of PERSONIFICATION-WITH-METONYMY to examine the agency attributed to AI technologies in discourse. Metonymy is understood as a mapping of a conceptual source domain onto a target domain, where both the source and the target are part of the same functional domain and "linked by a pragmatic function, so that the target is mentally activated" (Barcelona, 2011, p. 52). Thus, although metaphor and metonymy may often be interpreted to overlap, metonymy entails a contiguity relationship between domains, which are parts or aspects of the same "top-level" domain. Metaphor, on the other hand, entails an analogy-based relationship between two domains, which do not share a common top-level domain. PERSONIFICATION-WITH-METONYMY can be considered a type of personification, the phenomenon of assigning human-like traits to non-humans, which includes metonymy. While personification can be interpreted as metaphorical, metonymical, or its own category (Lakoff & Johnson, 2003), it is beyond the scope of this chapter to discuss types of and relations between metaphor, metonymy, and personification. For our purposes, we consider PERSONIFICATION-WITH-METONYMY a subtype of personification (Dorst, 2011; Dorst, Mulder & Steen, 2011) where the personified entity is based on metonymy. For example, in the phrase "AI *has contributed* to countering the current COVID-19 pandemic" (AI Watch, 2021a, p. 10), "AI" is a personified agent and also a metonymy, as it is used to represent *the use of AI* or *the development of AI*. In other words, it is not *actual AI* that is performing the act of contributing to countering a pandemic, but the actions taken by humans in using, developing, or funding AI initiatives, and so on. In sum, metaphors, metonymy, and personification can function as ways of studying how AI is presented and perceived.

We examined a corpus of all 20 reports (nearly 1,300 pages of English text) prepared by AI Watch, the European Commission's knowledge service on AI, published in the years 2020–2021. AI Watch represents a body of experts monitoring "industrial, technological and research capacity, policy initiatives in the Member States, uptake and technical developments of Artificial Intelligence and its impact in the economy, society and public services".[2] To this

[2] https://ai-watch.ec.europa.eu/about_en.

end, the AI Watch initiative, which is part of the European Commission's Joint Research Centre, serves as an information bridge between researchers and policymakers. The topics cover timely themes related to AI systems in Europe and the world from the use of AI in the public sector to national AI strategies to the definition of AI. The reports produced by AI Watch are an impactful source of language in that they are situated at the intersection of expert and policy narratives.

Framed as a classification task, metaphor detection can be performed either as human or automated annotation. As a manual annotation task, metaphors are often identified using the metaphor identification procedure (MIP) (Pragglejaz Group, 2007) or its close relative MIPVU (Steen et al., 2010), which entails annotation of each word in the text as metaphorical or not based on dictionary definitions, followed by inter-annotator agreement scoring. To ease manual labor, MIP annotation can be performed by selecting potentially metaphorical sentences as working hypotheses and applying MIP only to those sentences. As a computational task, large language models, often based on Google's BERT, have been employed for metaphor detection, typically taking the framework of CMT as a starting point, with some of the models incorporating MIP into their pipeline, such as in Choi et al. (2021).

To find instances of PERSONIFICATION-WITH-METONYMY in text, we conducted dependency parsing using *spaCy* (Honnibal et al., 2020), a free, open-source library for advanced NLP in Python. Dependency parsing allows querying sentences, which contain a relevant term, such as "AI", in the syntactic function of a subject. We analyzed the ways AI technologies are personified through the use of active verbs, such as in "AI is *bringing* many opportunities to the health system by means of data-driven medicine that can improve prevention, prediction and monitoring" (AI Watch, 2021a, p. 136). In the AI Watch reports, we found that "AI" performs many agentic actions, such as *providing*, *improving*, *helping*, *creating*, *supporting*, *achieving*, *protecting*, and *threatening*. An agentic AI operates across diverse sectors of society, including the labor market, healthcare, mobility, sustainability, and cybersecurity.

Metaphorically, the AI Watch reports we examined frequently characterize AI as a force and its uptake as a race. Following CMT, these metaphorical frames can be formalized as the CMs AI UPTAKE IS A RACE and AI IS A FORCE, with linguistic realizations such as "the EU can reinforce its position as a top AI player at global level" (AI Watch, 2021a, p. 7) or "governments are supporting human capacity building in AI and aim to prepare for the labour market transformations brought about by AI technologies" (AI Watch, 2021a, p. 11). Furthermore, the reports prolifically use the common metaphor of "trustworthy" AI, which also attributes agentic characteristics to the technology. In this vein, we could find several CMs entailing the personification of AI, such as AI IS AN ASSISTANT: "citizens send in challenges [...] on what they see as important societal challenges in their life *where AI could help*" (AI Watch, 2021a, p. 28). PERSONIFICATION-WITH-METONYMY and metaphorical language overlap and intersect, giving AI several roles in relation to humans. On one hand, AI is framed as a transformative force, which causes societal changes and necessitates reactionary measures from policymakers. Simultaneously, AI is presented as a facilitator of economic growth, such as when it "brings opportunities" and "could benefit governments, such as policy making, internal management, and improving public service delivery" (AI Watch, 2021b, p. 7).

While language needs space to develop alongside emerging multidisciplinary efforts around AI systems, AI represents a special case in the study of agency and personification. As a highly specialized field of expertise, AI not only has jargon – be it hyperbolic, metaphorical, or misleading – but also a long-standing tradition of personification, which stems from

its nature as a technology developed to mimic human actions. However, given their ubiquity, it is crucial to bring attention to present (and past) ways of conceptualizing and narrating AI systems. Whether we embrace the notion of AI as a friendly collaborator, a transformational force, a mystical black box, a combination, or none of these, warrants wider discussion. The personification of AI often leads to or stems from the concept of AGI as the ultimate goal for the field of AI and fear for the rest of society. Reorienting discussion away from AGI, which at present appears a faraway scenario and a concept irrelevant and misleading to any meaningful nature of AI, allows us to focus on the far-reaching effects of ANI already in ubiquitous use.

DELIBERATIONS AND STAKEHOLDER AGENCY IN RESPONSES TO THE AI ACT

The European Commission (EC) recently introduced the Artificial Intelligence Act (or the AI Act) as a proposal to establish harmonized rules for the regulation of AI technologies developed, placed, and used in the European Union (EU) market (EC AI Act 2021). It is based on "EU values and fundamental rights" and was prepared in response to calls for legislative action to ensure a well-functioning internal market for AI systems seeking to address both the benefits and risks of AI, without unduly constraining technological development or costs of developing AI solutions. To this end, it undertakes a "*proportionate horizontal* and *risk-based regulatory approach* to AI, based on a robust and flexible legal framework" (EC AI Act 2021). Since the introduction of this proposal, there has been a good deal of debate around the efficacy and practical feasibility of the proposal among AI researchers, industry practitioners, and policy experts, while consultations and deliberations among EU member states are continuing to shape its final formulation and approaches for compliance as an EU-wide regulatory framework in the near future. Much of the debate has centered around the very definition of "artificial intelligence" and what constitutes high versus low risk for harmful AI practices. Some of this debate has occurred in news media articles, policy briefs, and presentations.

The EC initiated consultations and solicited feedback on the proposal from organizations and practitioners in 2020 and 2021. We are examining the nature of these deliberations from respondents using both qualitative and quantitative analysis. Here we share some recurring themes and issues emerging through some responses during the consultation phase. For example, one organization pointed out that "the concept of Artificial Intelligence is still not well defined and as such we oppose regulating it, as there may be completely different understandings of AI by various actors on the market" (EC Consultations, 2020). Another suggested that while some high-risk applications may need strict regulation, defining them is a significant challenge.

> The application depends on the domain, but even more on the use case and the data - making the right level of granularity challenging. E.g., there are HR related AI applications which can be high risk for discrimination – but a chatbot assistant that answers questions on open positions would hardly be one. (EC Consultations, 2020)

In the AI Act, the EC proposed a single "future-proof definition" of AI in a supplementary annex. It devised a methodology for defining "high-risk" AI as systems that pose "significant risks to the health and safety or fundamental rights of persons" in several sensitive application domains that must conform with requirements for trustworthy AI before entering the

EU market. The deliberations around the definitions of AI and differing notions of risk have remained fraught throughout this consultative process and debates on the proposal. These discourses indicate the linguistic ambiguity of how artificial intelligence as a technology, system, or set of practices is broadly understood, the implications of what is considered harmful or risky even within the same application domain, and the diverse stakes involved among wide-ranging actors in such a regulatory exercise. Ultimately, the debate is really about the language of AI, how it is perceived or understood by different parties, and how the AI Act may empower or disempower their agency in creating, deploying, or offering AI services.

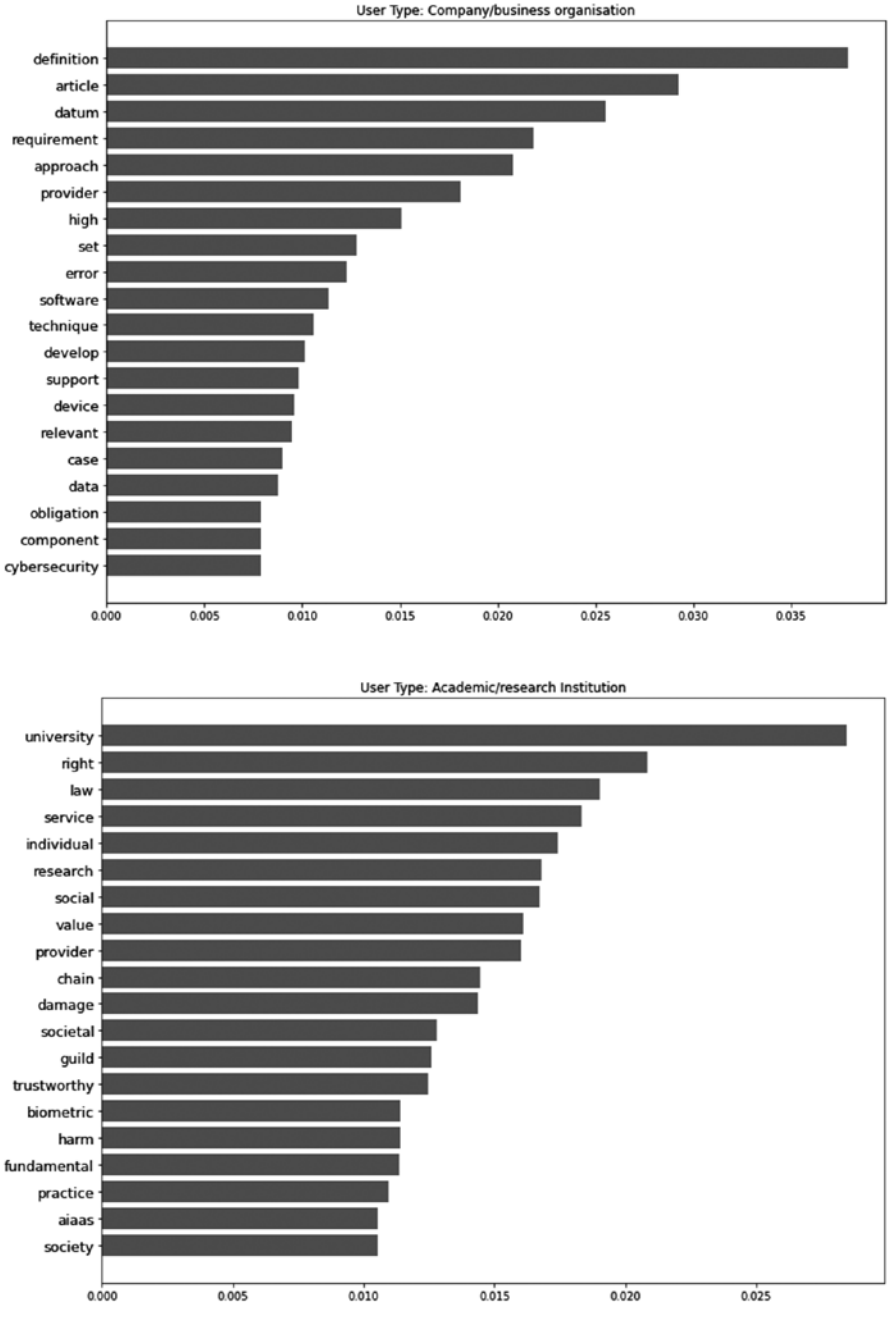
We examined English-language responses to the AI Act (approx. all 300 feedback documents on the EC online portal from April 26–August 6, 2021) through keyword extraction and n-gram ranking using spaCy and KeyBERT (Grootendorst, 2020) to explore preliminary implications of engaging multi-stakeholder feedback and hypotheses for further study. Industry actors and business associations are the most prominent types of organizations in the set of feedback documents, followed by non-governmental organizations (NGOs) and academic institutions. While this suggests that the feedback submitted in the EC online portal provides merely a snapshot of ongoing deliberations around the AI Act, it also highlights the prominence of industry narratives. Active engagement on the part of the types of stakeholders represented in the feedback documents is a realization of agency often available to large and established organizations (vs. civil society organizations or actors), which makes the critical examination of their deliberative language compelling.

Our initial data exploration using keyword ranking of normalized unigram frequencies shows the differences in potential themes and concerns raised by stakeholders (Figure 19.1). Companies and business organizations seem to raise issues around the definition of AI in the proposed regulations, the requirements, and obligations of compliance as well as implications for AI software development. Given their positions as service providers, manufacturers, and developers of AI-based tools, industry stakeholders are faced with challenges of matching regulatory compliance with financial aims, much of which boils down to the definition of AI operationalized in the AI Act. In contrast to industry actors, academic and research institutions seem to raise societal issues regarding the trustworthiness of AI systems, possible damage to individuals and diverse social groups, and human rights and democratic values.

The conceptualizations and narratives of AI, discussed throughout this chapter, are not absent from regulatory deliberations. We can examine the AI Act as a turn or shift in the wider narrative of AI, requiring multiple parties to react and possibly shaping the direction of future discourse. Among many narrative shifts, the AI Act represents one such event, necessitating multi-stakeholder action, which may lead to both overt and covert reimagination, reformulation, and reorientation processes in the language of AI.

CONCLUSIONS: REFRAMING AGENCY IN AI SYSTEMS

Artificial intelligence is "no longer an engineering discipline" (Dignum, 2020). Multidisciplinary efforts are needed, and as the field takes on new forms and disciplinary identities, language needs to function as a meaningful bridge between domains of knowledge. Multidisciplinarity in AI introduces a range of challenges as new perspectives come into play. The encounters and intersections between discursive practices among various actors inform how those practices feed into the critical understanding of both the essence of AI as well as its implications



Source: Visualization by Shintaro Sakai, summer researcher at Aalto University, 2022.

Figure 19.1 *Salient unigrams in the AI Act feedback documents submitted by academic and other research institutions*

across diverse sectors of society. Beyond understanding, however, critically focusing on the use of language can facilitate algorithmic literacy among researchers, practitioners, and citizens. Following the argument of Rehak (2021), policymakers and regulators would benefit from researchers and science journalists focusing on "constructive" ways of using language around AI, such as avoiding exaggerated anthropomorphisms pervasive in popular science and culture, hyperbolic terminology (Bones et al., 2021), and technical terms such as *communication* and *recognition*, which often translate into popularized metaphors. Beyond anthropomorphic terminology, such as *learning*, *training*, or *memory*, AI technologies are often presented as having (human-like) agency through personification. Agency is clearly embedded in human actions across the development, deployment, regulation, and user experience stages of an AI system's life cycle. However, as we have demonstrated in this chapter, in popular AI discourses, the concept of *action* by machines is often conflated with their observable *behavior* (Searle, 1992), perceived as autonomy, agency, or sentience in AI, and vice versa. That is, AI systems are often framed as agents operating independently in society, requiring humans to adjust to their emergence and form trust relationships with them, almost as if we lived in a world where AGI systems were a reality. Addressing this quandary, Rehak (2021) calls for discarding terms like "agency" and "autonomy" entirely from how we speak about artificial intelligence, as they can be both inaccurate and misleading; but more importantly, we need to shift our focus to human agency implicit in the design, development, operation, and governance of AI systems. This reframing of agency mitigates the sociotechnical blindness implicit in such discourses, while emphasizing a greater awareness of sociocultural contexts, power structures, and political practices that shape the way AI systems emerge, function, and proliferate in society.

REFERENCES

Adam, A. (1995). Artificial intelligence and women's knowledge: What can feminist epistemologies tell us? *Women's Studies International Forum*, 18(4), 407–415.

AI Now Institute. (2018). AI now report 2018. *Tech. Rep.* [Online] www.ainowinstitute.org.

AI Watch. (2021a). National strategies on Artificial Intelligence: A European perspective. JRC-OECD report.

AI Watch. (2021b). Artificial Intelligence for the public sector. Report of the "3rd Peer Learning Workshop on the use and impact of AI in public services", 24 June 2021. JRC Conference and Workshop Report.

Alexander, P. A., Schallert, D. L., & Reynolds, R. E. (2009). What is learning anyway? A topographical perspective considered. *Educational Psychologist*, 44(3), 176–192.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989.

Barcelona, A. (2011). Reviewing the properties and prototype structure of metonymy. In Réka Benczes, Antonio Barcelona, and Francisco José Ruiz de Mendoza Ibáñez (Eds), *Defining metonymy in cognitive linguistics: Towards a consensus view*, 7–57. John Benjamins Publishing Company.

Bones, H., Ford, S., Hendery, R., Richards, K., & Swist, T. (2021). In the frame: The language of AI. *Philosophy and Technology*, 34(1), 23–44.

Bunz, M., & Braghieri, M. (2022). The AI doctor will see you now: Assessing the framing of AI in news coverage. *AI & Society*, 37, 9–22.

Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Palgrave Macmillan.

Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D., & Lee, J. (2021). MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics:*

Human Language Technologies, pages 1763–1773. Online. Association for Computational Linguistics.
Crawford, K. (2021). *The atlas of AI.* Yale University Press.
Dewandre, N., & Gulyás, O. (2018). Sensitive economic personae and functional human beings: A critical metaphor analysis of EU policy documents between 1985 and 2014. *Journal of Language and Politics*, 17(6), 831–857.
Dignum, V. (2020). AI is multidisciplinary. *AI Matters*, 5(4) (December 2019), 18–21.
Dorst, L. (2011). Personification in discourse: Linguistic forms, conceptual structures and communicative functions. *Language and Literature*, 20(2).
Dorst, A. G., Mulder, G., & Steen, G. J. (2011). Recognition of personifications in fiction by non-expert readers. *Metaphor and the Social World*, 1(2), 174–200.
Dreyfus, H. L. (1972). *What computers can't do.* New York, NY: Harper & Row.
Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. *Free Press.*
Drobotowicz, K., Kauppinen, M., & Kujala, S. (2021). Trustworthy AI services in the public sector: What are citizens saying about it? In F. Dalpiaz and P. Spoletini (eds.), *Requirements engineering: Foundation for software quality.* REFSQ.
Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1).
Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C. R., . . . others. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management,* 57.
Elish, M. C., & boyd, d. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 85(1), 57–80.
European Commission. (2020). Public consultations on Artificial intelligence – ethical and legal requirements, 20 February 2020 – 14 June 2020. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements_en.
European Commission. (2021). Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, COM/2021/206 final, Brussels, 21.4.2021. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN.
Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7(10).
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., . . . others. (2018). Ai4people—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.
Floridi, L. (2020). Artificial Intelligence as a public service: Learning from Amsterdam and Helsinki. *Philosophy & Technology*, 33, 541–546.
Grootendorst, M. (2020). KeyBERT: Minimal keyword extraction with BERT. *Zenodo.org.* DOI: 10.5281/zenodo.4461265.
Haataja, M., van de Fliert, L., & Rautio, P. (2020). Public AI registers: Realising AI transparency and civic participation in government use of AI. Whitepaper Version 1.0, *Saidot*, September.
Hermann, E. (2021). Artificial intelligence and mass personalization of communication content— An ethical and literacy perspective. *New Media & Society.* 24(5), 1258–1277.
Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python. *Zenodo.org.* DOI: 10.5281/zenodo.1212303.
Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. *Minds & Machines*, 27, 575–590.
Kövecses, Z. (2002). *Metaphor: A practical introduction*. New York: Oxford University Press.
Kövecses, Z. (2020). An extended view of conceptual metaphor theory. *Review of Cognitive Linguistics*, 18(1), 112–130.
Lakoff, G., Espenson, J., & Schwartz, A. (1991). *Master metaphor list.* Berkeley, CA: Cognitive Linguistics Group University of California. (Technical report).
Lakoff, G., & Johnson, M. (2003). *Metaphors we live by.* The University of Chicago Press.
Lavin, M. (2021). Why digital humanists should emphasize situated data over Capta. *Digital Humanities Quarterly*, 15(2).

Lemoine, B. (2022). Is LaMDA sentient? — An interview. *Medium*, June 11. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16.

Minsky, M. (1975). A framework for representing knowledge. In Pat Winston (ed.), *The psychology of computer vision*. New York: McGraw Hill, 211–277.

Ossewaarde, M., & Gulenc, E. (2020). National varieties of Artificial Intelligence discourses: Myth, utopianism, and solutionism in West European policy expectations. *Computer*, 53(11), 53–61.

Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1–39.

Rehak, R. (2021). The language labyrinth: Constructive critique on the terminology used in the AI discourse. In P. Verdegem (ed.), *AI for everyone?*. London: University of Westminster Press.

Sandis, C., & Harper, R. (2017). We don't want AI that can understand us – we'd only end up arguing. *The Conversation*, August 21. https://theconversation.com/we-dont-want-ai-that-can-understand-us-wed-only-end-up-arguing-82338.

Sawhney, N. (2022). Contestations in urban mobility: Rights, risks, and responsibilities for Urban AI. *AI & Society*.

Searle, J. R. (1992). *The rediscovery of mind*. Cambridge, MA: MIT Press.

Sparkes, M. (2022). Has Google's LaMDA artificial intelligence really achieved sentience? *New Scientist*, June 13. https://www.newscientist.com/article/2323905-has-googles-lamda-artificial-intelligence-really-achieved-sentience/.

Steen, G., Dorst, L., Herrmann, J., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins Publishing Company.

Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*, LIX(236), October, 433–460.

Wyatt, S. (2021). Metaphors in critical Internet and digital media studies. *New Media & Society*, 23(2), 406–416.