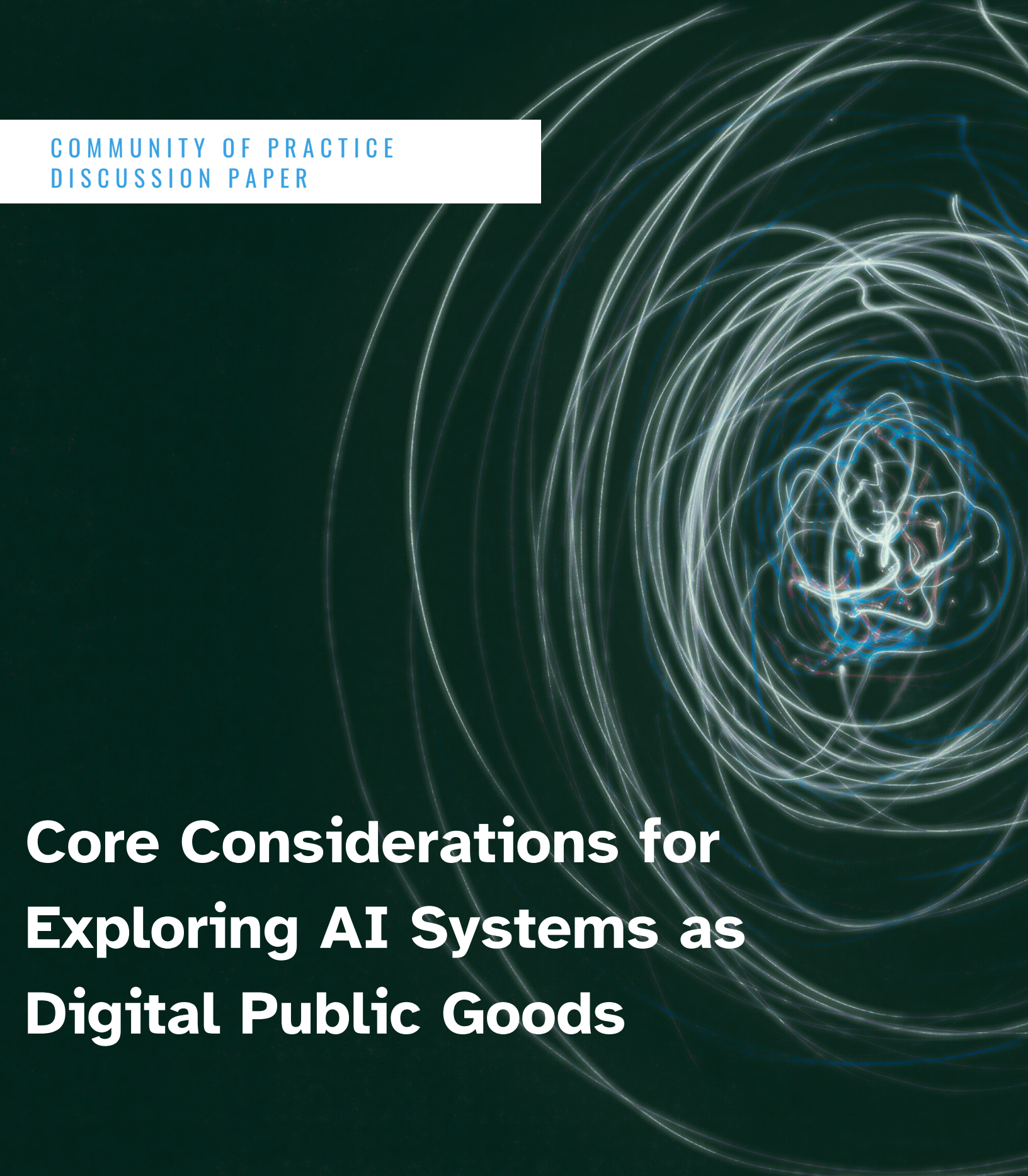
COMMUNITY OF PRACTICE
DISCUSSION PAPER



Core Considerations for Exploring AI Systems as Digital Public Goods

Digital Public Goods Alliance

unicef
for every child

COMMUNITY OF PRACTICE
DISCUSSION PAPER

Core Considerations for Exploring AI Systems as Digital Public Goods

Established in 2019, the Digital Public Goods Alliance is a multi-stakeholder initiative with a mission to accelerate the attainment of the sustainable development goals in low- and middle-income countries by facilitating the discovery, development, use of, and investment in digital public goods. Digital public goods are open-source software, open data, open AI systems, and open content collections that adhere to privacy and other applicable laws and best practices, do no harm, and help attain the SDGs. To learn more, visit digitalpublicgoods.net or contact hello@digitalpublicgoods.net

Cover photo credit: Gertrūda Valasevičiūtė

unicef
for every child

COMMUNITY OF PRACTICE
DISCUSSION PAPER

Core Considerations for Exploring AI Systems as Digital Public Goods

Table of Contents

1. Introduction 1
2. Understanding the "open AI" and "responsible AI" debate and focus of the AI community of practice 2
3. Relevant concepts and core questions 3
3.1 Understanding "AI systems" and their features 3
3.2 Describing responsible AI 6
3.3 The data layer - considerations drawn from the DPGA's Open Data CoP and beyond 7
4. Tackling the open AI dimension 8
5. Tackling the responsible AI dimension 9
Acknowledgements 11
Appendix - Selected additional resources 12

1. Introduction

The unprecedented advancement of artificial intelligence (AI) technologies was catalysed by a strong emphasis on open collaboration and sharing. In fact, the speed of innovation we have seen in recent years is powered by making research, data, and models openly accessible. As AI technologies become more prevalent, there is an urgent need to ensure they benefit society. AI can serve as a powerful tool for good, particularly in the realm of international development. Recognising the timeliness of addressing these crucial issues, the Digital Public Goods Alliance (DPGA) and UNICEF have joined forces to form a community of practice that brings together experts from diverse sectors and geographies. Together, they aim to explore the intersection of ethical AI and open source, and their efforts will inform essential changes to the Digital Public Good (DPG) Standard, which plays a pivotal role in determining what constitutes AI systems as digital public goods.

The urgency surrounding the ethical use of AI arises from its potential implications across various aspects of society. As AI systems become increasingly sophisticated and integrated into our daily

lives, concerns over privacy, bias, and accountability and related risks both short- and long-term have come to the forefront of discussions on AI development and governance. To ensure that AI is harnessed responsibly, it is vital to establish reliable governance mechanisms. Given the uneven speed of technology development and regulatory frameworks development, the perspective now shifts towards combining upstream (managing the innovation process) and downstream governance approaches (managing the risks of technology development)[1]. These include ethical frameworks for the development and deployment of AI. Recognising the potential for AI systems to address complex global challenges, bridge digital divides, and advance sustainable development goals, solutions must be found.

By bringing together a wide range of perspectives, the community of practice (CoP) aims to examine the intersection of ethical AI and open source. Through collaboration and knowledge sharing, this group will delve into the complex considerations associated with AI development, its democratisation (i.e open-sourcing underlying data and models), and identify strategies for how responsible open-source AI systems can be digital public goods. While the current definition of DPGs includes "AI models", the CoP is taking a more comprehensive approach by focusing on "AI systems" instead, including the data layer, output layer, and human-machine interaction layer. The work of the CoP will serve as a guide for updating the DPG Standard, a crucial benchmark that defines the qualities and characteristics of digital public goods.

Given the vast questions that surface in conversations of responsible AI and open source, the DPGA and UNICEF recognise that various open questions will remain unanswered through this work. However, it is crucial to take action in order to establish a framework that guides developers, funders, and implementers in their work on open-source AI, harnesses its potential for positive impact, and ensures that AI digital public goods contribute to a more inclusive and sustainable future for the whole world. The following report provides insight into the initial progress made by the CoP while establishing the frameworks of understanding it will work within and what it will strive to achieve.

2. Understanding the "open AI" and "responsible AI" debate and focus of the AI community of practice

The debate between open and responsible AI is often framed as a trade-off between two critical goals: on the one hand, open-source AI systems as freely accessible, usable, modifiable, and distributable digital artefacts that allow for adaptation and adoption that can can bring numerous benefits including fostering further innovation. On the other hand, the need for responsible and ethical AI that is designed and implemented in a way to benefit society as a whole which can be a challenging task when fully open.

Although both debates have public value and societal benefits at their core, recent discussions and cases have challenged the idea that radical openness does in fact benefit society. For example, open indigenous language datasets have brought into question concepts of ownership and the

[1] "Technology Governance", OECD, https://www.oecd.org/sti/science-technology-innovation-outlook/technology-governance/

perpetuation of existing exploitive relationships and cultural distortion due to an imbalance of available resources for AI system development, scaling, and use. Instead, there is a need to build equitable AI infrastructures, including within disadvantaged and low resource settings.

While openness and responsibility of AI systems are at the centre of their own separate debates, the DPGA's community of practice, co-chaired by UNICEF Office of Innovation, is exploring the intersection of these two debates. Exploration includes, but is not limited to, experimental licences for datasets and models to ensure benefits flow back to communities and those traditionally disadvantaged, as well as tools, frameworks, and best practices to ensure and document where AI is developed and deployed responsibly. As a starting point, the CoP will work within the framework of the DPG Standard to evolve as a mechanism that can assess AI systems as digital public goods.

In addition to preventing harm, there is a need to explore how to navigate ambiguities stemming from the current definition of "open" in relation to data extractivism, data colonialism, and the economics of open data with regard to AI development and its required resources such as computing power and access to finance. Recognising that these questions are part of a larger discussion occurring within open-source communities, this work will collaborate with and closely follow other worksteams, such as Open Knowledge Foundation's recently launched initiative to redefine the open definition to meet today's challenges. The DPGA and UNICEF are also working in close collaboration with the Open Source Initiative's effort to specifically define open-source AI.

Additionally, with recognition that many developers rely on freely available datasets from sources such as Google Earth to build their AI products, the CoP will make recommendations on the extent to which datasets need to be open and which licences will be required for AI training and testing datasets for an AI system to be considered under the DPG Standard. In doing so, we must consider the tradeoff between limiting the (public) value AI systems can generate without openly available datasets versus a limited number of AI systems that are completely freely accessible, usable, modifiable, and distributable as digital public goods.

While recommendations for legislation, policies, and other instruments to govern AI systems are needed, this CoP will focus on how and to what extent the principles of responsible AI should be reflected in the DPG Standard to ensure that only AI systems that adhere to trustworthy and responsible AI principles, besides questions around open licensing and data privacy, will be considered DPGs.

3. Relevant concepts and core questions

3.1 Understanding "AI systems" and their features

While there is an abundance of work covering the topic of responsible AI (including "ethical" AI and "trustworthy" AI)[2], there are few resources that define the features of open-source AI and

[2] "Frameworks and Tools on Responsible AI", Global Index on Responsible AI, contains a list of existing tools and frameworks seeking to advance the ethical, responsible or trustworthy development and use of AI.https://www.responsibleaiindex.org/frameworks-and-tools

mechanisms for opening up AI systems. As such, there is a need to find a common understanding. This section aims to outline various perspectives on what is included within an AI system. Below outlines what the CoP considers to be an AI system:

> *"AI systems are machine-based systems designed to operate with varying levels of autonomy that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments or generate outputs, such as text, images, or sounds."*[3]

Various stakeholders have developed representations of AI systems' components. Github published a paper featuring a comprehensive representation of an AI system from a technical perspective which they refer to as an AI Software Stack. However, the CoP is exploring AI systems from a socio-technical perspective. The layered approach in figure 1 is a simplified overview of the different AI system components from a socio-technical perspective.

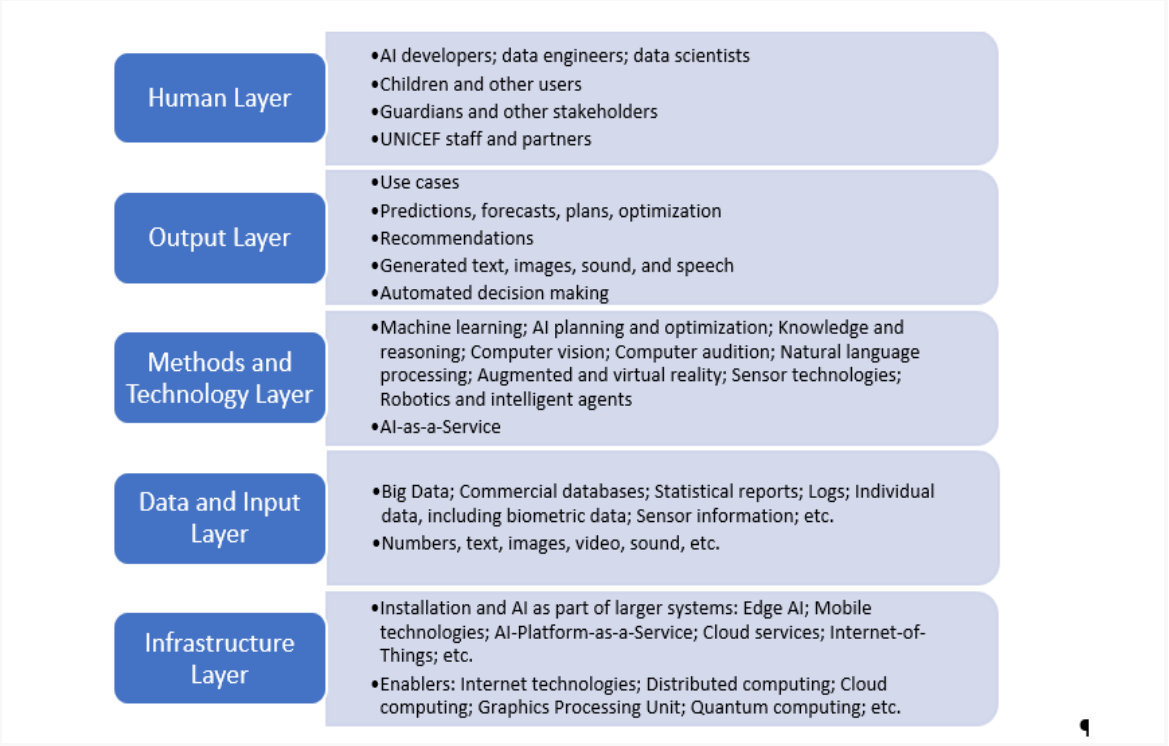


Figure1: AI systems' layers from a socio-technological perspective (Source: UNICEF)

Another representation of an AI system's layers was created by the OECD Working Group on AI Classification. The key dimensions are as follows:

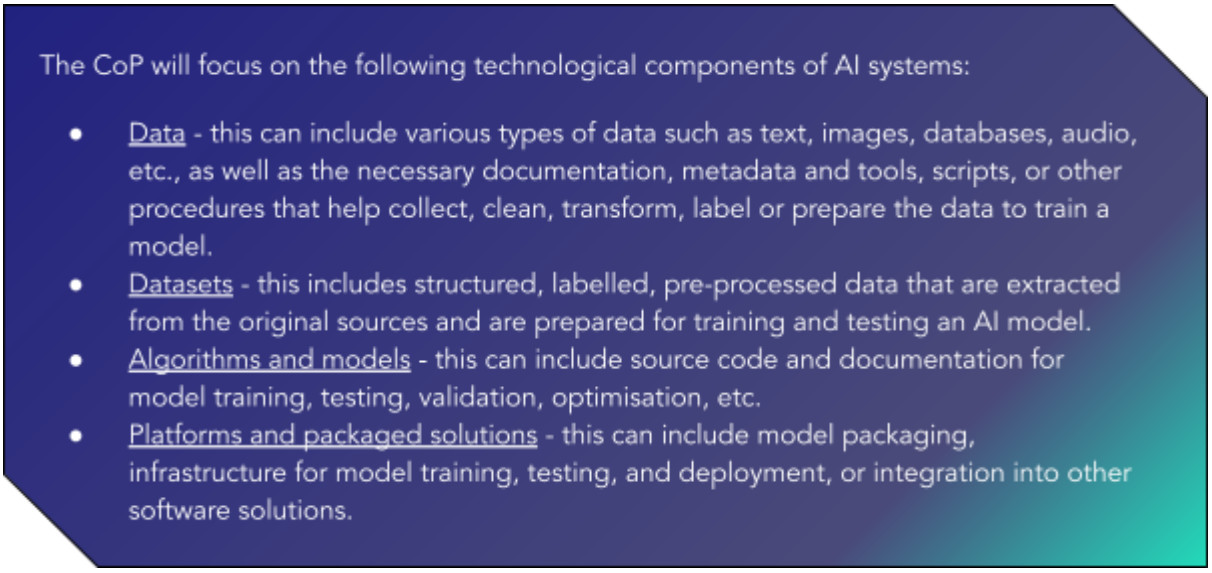
- People and Planet: users and impacted stakeholders, the application's optionality and how it impacts human rights, the environment, well-being, society, and the world of work.
- Economic Context: the economic and sectoral environment in which an applied AI system is implemented.

[3] This definition aligns with OECD guidelines on AI ("Recommendation of the Council on Artificial Intelligence", OECD Publishing, Paris 2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449) and the definition adopted by UNESCO: "AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments." (UNESCO 2022).

- Data and Input: the provenance of data and inputs, machine and/or human collection method, data structure and format, and data properties.
- AI Model: technical type, how the model is built (using expert knowledge, machine learning or both) and how the model is used (for what objectives and using what performance measures).
- Task and Output: tasks the system performs, e.g. personalisation, recognition, forecasting or goal-driven optimisation; its outputs; and the resulting action(s) that influence the overall context.

There are many similarities between the two frameworks above and both models provide a helpful characterization of the components we will discuss in the CoP. The OECD model offers the possibility to explore "openness" in different sectors by providing a context-dependent perspective. For instance, openness in health, agriculture, and education may mean different things. There are more complex descriptions available, including approaches combining a layered model as described above with a description of the AI lifecycle to capture key technical aspects and the AI actors[4] involved at each stage (OECD 2019). For the purpose of the CoP, we believe that a simplified model can serve as the foundation needed to inform the discussion.

The debate on what constitutes open-source AI systems is still nascent, and there's a need to explore all of the layers mentioned above. Many of these considerations can only serve as a framing and backdrop of the CoP's discussion, while we focus on some key components which can be reflected in the DPG Standard. This includes the requirement of opening training and testing datasets and under which terms; licences of AI algorithms, models, and other related artefacts (data, code, content, etc); its integration on platforms, packaged solutions; and measures to ensure a do-no-harm by design approach.



Note: These are technical components we can expect from AI systems, but this list is not exhaustive. We also note that there are projects which develop AI systems from scratch, as well as solutions that rely on other

[4] According to the OECD, "AI actors are those who play an active role throughout the AI system lifecycle and can include organisations and individuals that deploy or operate AI" (OECD 2019). The stages of the AI system lifecycle include planning and design; collecting and processing data; building and using the model; verifying and validating; deployment; and operating and monitoring.

(pre-trained) models or build on top of those, as well as the use of packaged AI services (e.g. Azure Cognitive Services or Google Cloud AI services). The community of practice, however, intentionally limits the scope of inquiry to consider only the components that could be defined in the DPG Standard.

3.2 Describing responsible AI

The core work for the CoP will lie in exploring the intersection between "open" and "responsible" AI. The following understanding for "responsible AI" will be used:

> *Responsible AI encompasses a broader set of principles that address not only the ethical considerations of AI, such as moral values and principles[5] but also the social, legal, and economic implications of AI. Responsible AI ensures that AI systems are transparent, explainable, and accountable and that they are developed and deployed to promote diversity, equity[6], and inclusion.*

Within the framework of this understanding, the features of responsible AI systems include[7]:

1. Transparency: The AI system should be transparent regarding how it makes decisions, what data it uses, and how it processes that data. It should also be transparent about any biases or limitations that may exist.
2. Fairness: The AI system should be designed in a participatory and inclusive manner to avoid discrimination or bias against any particular group of people. It should include continuous user feedback loops, and be tested to ensure that it does not perpetuate or amplify any existing biases in the data.
3. Privacy and security: The AI system should be designed to protect the privacy and security of user data, as well as models and users. It should also comply with applicable data protection laws and regulations.
4. Accountability: The AI system should include clear accountability and responsibility for its decisions and actions. There should be clear lines of responsibility for the system's outcomes, transparent and easily discoverable rules according to which outputs are provided, and easily accessible mechanisms for addressing errors or unintended consequences.
5. Human oversight: The AI system should have appropriate levels of human oversight to ensure robustness and integrity and that it operates ethically and responsibly. This may include consistently monitoring its decision-making processes, verifying its outputs, and intervening when necessary.

[5] Ethical AI refers to the development and deployment of AI systems that align with moral values and principles. This includes ensuring that AI is used to promote the well-being of individuals and society, and that it does not cause harm or discriminate against any particular group. Ethical AI also involves ensuring that AI systems respect privacy, autonomy, and human rights.

[6] Equity as a universal objective must be defined according to the context an AI solution is operating in. Here's one example which identifies equity definitions in health to ensure quality before training or using AI: https://arxiv.org/abs/2203.05174

[7] The verification of self-declaration on an AI system being open and responsible is out-of-scope for this community of practice. We acknowledge that there is an incentive to self-declare AI systems as being "bias-free", "equitable", "fair", and so on both from legal perspectives and from a market positioning perspective. We acknowledge that a process of public verification (such as a conformance program that the Linux Foundation uses to govern its open source programs) would be preferable to self-declaration and leave this for future work by this community of practice and other organisations developing AI standards.

6. Sustainability: The AI system should be designed to minimise its environmental impact[8] by using energy-efficient hardware and reducing its carbon footprint.
7. Risk mitigation: The AI system should include easy-to-access options for users to challenge decisions, request redress and enforcement mechanisms to ensure adherence to these outlined principles.

Note: Because the DPG Standard only focuses on core artefacts, *and their design, rather than implementations, points 5 and 6 will be outside the scope of the CoP. Which of these principles and how they can be reflected in the DPG Standard ultimately also depends on what is going to be featured in the DPG Registry. This could include both only AI models and complex AI solutions, and the compliance requirements for each of them might differ. Lastly, we also have to consider trade-offs between these principles, as the goal of one might compromise the other.*

The aforementioned features are informed by a number of sources, including the UNESCO recommendations on the ethics of AI, adopted by 193 countries. They provide a framework for, and starting point to elaborate on, governance requirements product owners need to adhere to and a backdrop for the responsible AI discussion within the CoP.

3.3 The data layer - considerations drawn from the DPGA's Open Data CoP and beyond

We cannot speak about open AI systems without speaking about data, as conflicts often arise from the details of open data access and licensing requirements.

Managing data as a fully non-excludable resource with no limitations on access, such as through an open data licence, removes barriers and maximises transparency and reusability[9]. The challenge, then, is how to ensure that AI is made as transparent, accountable, and as openly accessible as possible while safeguarding individual and collective rights and limiting access to data with the potential to do harm. Some datasets without the potential for accidental or intentional misuse could be managed as a digital public good. However, for datasets that contain sensitive information or otherwise carry the potential for misuse, access must be limited and managed[10,11]. As an exploration by Open Future outlines, we need more fine-grained control mechanisms for opening up data. In the same vein, we need to consider the tension that increased data access[12] creates for data subjects and data producers, especially from marginalised groups and disadvantaged geographies and their ability to realise collective benefits, amongst others[13].

[8] Currently, this principle might evoke tensions between offering life-saving interventions and minimising the environmental footprint. In addition, more elaboration and research regarding the measurement of an AI system's impact on sustainability is needed.
[9] "AI Commons: Filling the governance vacuum related to the use of information commons for AI training", Open Future, 2023, https://openfuture.eu/wp-content/uploads/2023/01/ai-commons-report.pdf. See this report published by Open Future for a detailed assessment of the challenges of open data and open content licensing for AI training data, and the need for commons-based governance models for AI training datasets.
[10] "Exploring Data as and in Service of the Public Good", Digital Public Goods Alliance and UNICEF 2023, https://digitalpublicgoods.net/PublicGoodDataReport.pdf
[11] For example, the GDPR defines health data as a special category, which requires additional safeguards. This also extends to AI training data.
[12] Note that tensions and challenges might differ depending on how datasets were compiled: through scraping open content on the internet, data sharing and data shared through purchases. Open Future offers a comprehensive take on the challenges of the first, including for dataset governance - see footnote 9.
[13] "CARE Principles for Indigenous Data Governance", Global Indigenous Data Alliance (GIDA), https://www.gida-global.org/care

As long as there is no formal definition of open-source AI, the DPGA requires open data licensing of training and testing datasets for AI systems as a DPG, bearing in mind reproducibility and platform independence (especially vis-à-vis freely provided data). Consequently, many AI products might not be recognised as a DPG at the moment, given their training and testing data sources or the need to adhere to privacy law and safety considerations.

Relevant core questions for the CoP to explore are:

- Do AI training and testing datasets need to be open for AI systems to be considered a DPG?
  - What are the consequences of unavailable datasets, and how does it impact the value an open AI system can deliver for developers and society?
  - What's the trade-off between using free (e.g. Google Earth) and open datasets (e.g. Mozilla Common Voice)?
- If so, which would be the open data licences or requirements needed for datasets to be considered "open"?
- If not, what other transparency mechanisms regarding training and testing datasets need to be put in place to ensure "do no harm by design" (e.g. biases)?
- Can we think of any solutions to make AI training and testing datasets openly available while safeguarding privacy and security?

4. Tackling the open AI dimension

Open Future has been exploring questions related to openness and AI. They map mechanisms for opening up AI systems at several levels of the AI ecosystem. These levels include:

- Regulation
- Licences
- Documentation and transparency
- (Participatory) Governance
- Access
- Resources/sustainability

Given the limited scope of the CoP, we will predominantly focus on the level of licences and documentation and transparency as shown in figure 2.

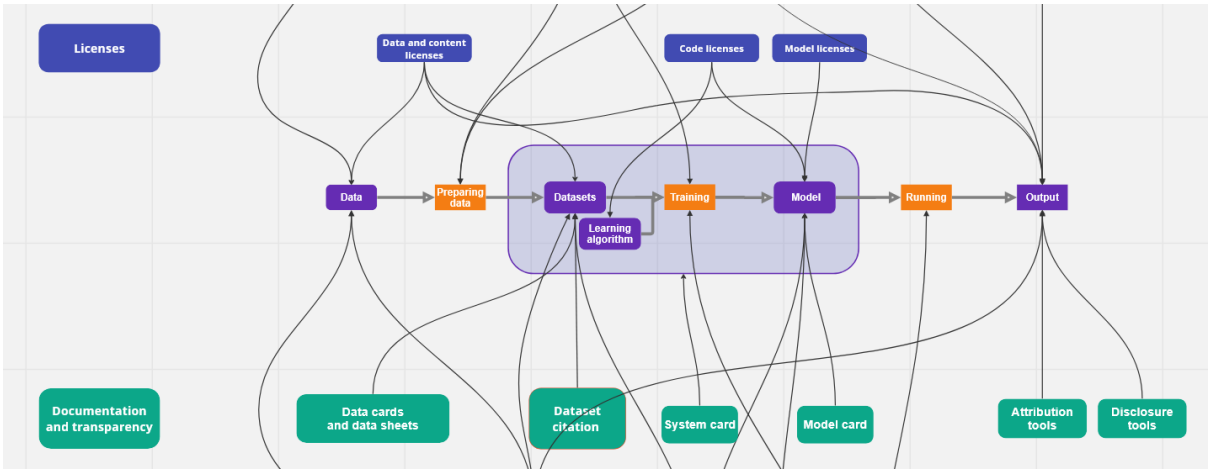


Figure 2: Licences and documentation & transparency layer mapped to the different AI systems' components (Source: Open Future)

Relevant core questions for the CoP to explore are:

- Which features constitute open responsible AI systems?
- How does this impact questions around documentation, licensing and access regarding AI systems as a DPG, taking into account the different components of an AI system?
- How can we measure/assess/document/monitor these defining aspects?
- Which of these aspects should be included in the DPG Standard and how?*

**Note: the aim is to build on the existing 9 indicators of the DPG Standard and discuss what specific aspects we can add under those rather than expanding or adding additional indicators.*

5. Tackling the responsible AI dimension

Considering the advantages and disadvantages of open source in AI, a dichotomy emerges between implementation-based harms and use-based harms. While responsible AI principles and tools to implement and monitor the deployment of AI systems offer remedies for implementation-based harms such as bias in training data, freedom of use is a fundamental component of open-source licensing. Most open-source developers don't have an overview of how their artefacts are used and don't consider themselves liable.

The CoP needs to explore measures for responsible AI, including how to provide incentives and motivation to instil responsible practices and norms for use in open-source AI communities to prevent harm and how to adopt novel technical approaches to monitor downstream use. At the same time, we need to acknowledge that this is not a black and white picture, where we either can prevent

all harms or none. Every measure will be incomplete and imperfect, but it's imperative to start developing and testing as many approaches as possible.[14]

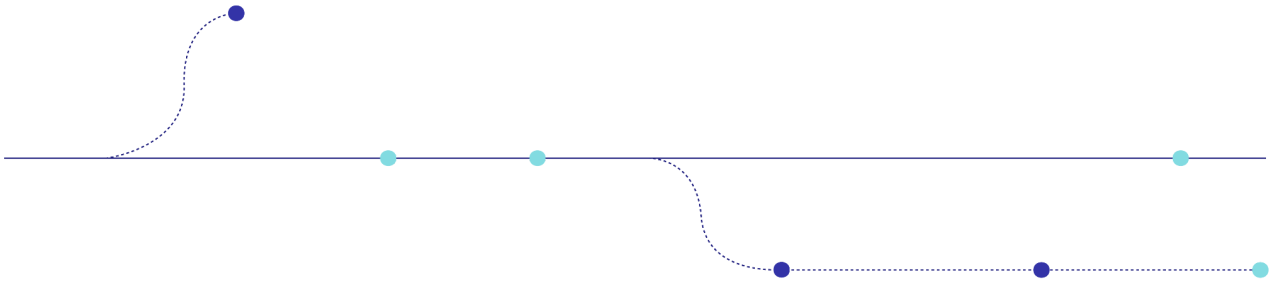
One initiative that tries to balance the requirements of open and responsible AI is the RAIL licences[15], combing open access and purpose limitation:

> *Responsible AI Licences (RAIL) empower developers to restrict the use of their AI technology to prevent irresponsible and harmful applications. These licences include behavioural-use clauses which grant permissions for specific use cases and/or limit certain use cases. If a licence permits derivative works, RAIL Licences also require the use of any downstream derivatives (including use, modification, redistribution, and repackaging) of the licensed artificial intelligence system must abide by the behavioural-use restrictions.*

However, RAIL doesn't consider itself open-source, and open-source enthusiasts criticise that RAIL licences limit the freedom to use a digital artefact without limitations[16]. Similarly, questions remain about incentivising the use of these licences and exercising control.

The core questions the CoP, thus, needs to answer are:

- Which areas of the responsible AI principles should we prioritise and focus on?
- How can we make them practical and enforce compliance?
- Are there other legal or licence considerations to take into account for other AI inputs and outputs (beyond datasets) that help instil compliance with responsible AI practices (e.g. RAIL)?
- How do we deal with the fact that AI systems constantly evolve, and any assessment or documentation is a static snapshot of a dynamic system?



[14] "Limits and Possibilities for "Ethical AI" in Open Source: A Study of Deepfakes", FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (Pages 2035–2046), June 2022, https://doi.org/10.1145/3531146.3533779
[15] "Behavioral Use Licensing for Responsible AI", FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (Pages 778–788), June 2022, https://doi.org/10.1145/3531146.3533143
[16] "4 freedoms", Wiki Open Source Ecology, https://wiki.opensourceecology.org/wiki/4_freedoms

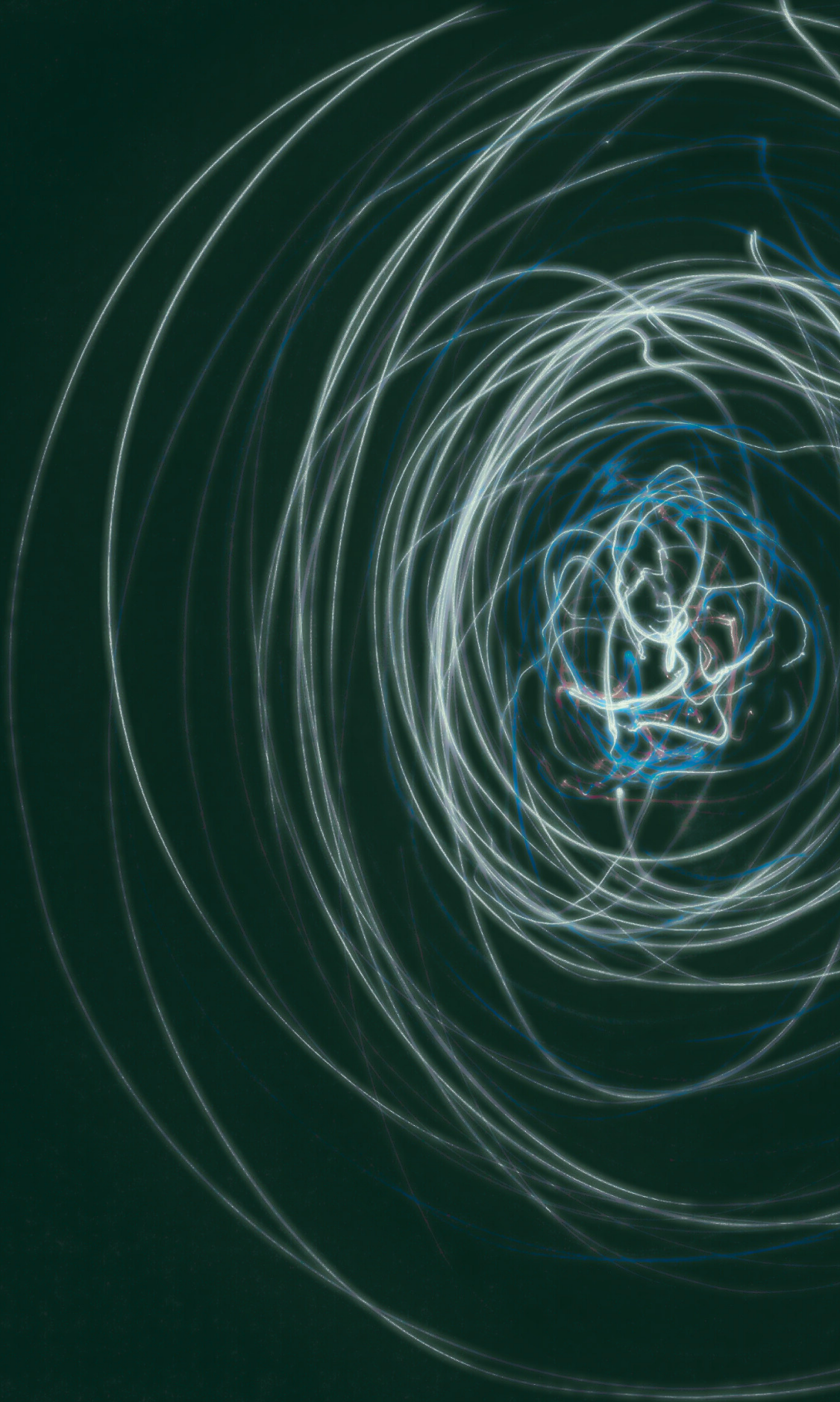
Acknowledgements

Communities of practice bring together experts from different institutions and networks, who meet to discuss how digital public goods can be deployed to address critical development needs and challenges. Many thanks to the people listed below for their participation and inputs that have helped shape and inform this discussion paper.

- Aaditeshwar Seth, GramVaani, IIT Delhi
- Abdoulaye Diack, Google
- Abraham Sam, Tony Blair Institute
- Alek Tarkowski, Open Future Foundation
- Asli P. Rhodin, Volvo
- Carlos Muñoz Ferrandis, Hugging Face
- Christian Resch, GIZ
- Darlington Akogo, Kara Agro and MinoHealth
- Daniel Brumund, GIZ
- Davide Storti, UNESCO
- Deb Bryant, Open Source Initiative
- Dina Machuve, Data Science Africa
- Emmanuel Letouze, Data Pop Alliance
- Filippo Pierozzi, United Nations
- Friederike von Franqué, Wikimedia Germany
- Govind Shivkumar, Omidyar Network
- Gustavo Fonseca Ribeiro, UNESCO
- Helani Galpaya, Wadhwani Institute for AI
- Ignatius Ezeani, Lancaster University
- Irina Mirkina, UNICEF
- Jaan Altosaar Li, One Fact Foundation
- Jameson Voisin, Digital Public Goods Alliance
- Jed Sundwall, Radiant Earth
- Jerry Kponyo, Kwame Nkrumah University of Science and Technology
- Kathleen Siminyu, Mozilla Foundation
- Keegan McBride, Oxford Internet Institute
- Lea Gimpel, Digital Public Goods Alliance
- Linda Bonyo, Lawyers Hub Africa
- Merl Chandana, LIRNEasia
- Michael Canares, Step Up Consulting
- Moritz Fromageot, UN
- Moses Thiga, Kabarak University
- Prajakta Kuwalekar, PeoplePlus AI
- Prateek Sibal, UNESCO
- Ravit Dotan, CAIR Lab
- Rebecca Ryakitimbo, Mozilla Foundation
- Ricardo Miron Torres, Digital Public Goods Alliance
- Romina Garrido, GobLab Universidad Adolfo Ibáñez
- Ruth Schmidt, GIZ
- Sarah Watson, Digital Public Goods Alliance
- Savvina Papadaki, Samsung
- Stefano Maffulli, Open Source Initiative
- Susan Aaronson, George Washington University
- Tanuj Bhojwani, PeoplePlus AI
- Tarunima Prabhakar, Tattle Civic Technologies
- Urvashi Aneja, Digital Futures Lab
- Zuzanna Warso, Open Future

Appendix - Selected additional resources

1. The European Commission's High-Level Expert Group on AI, which developed a set of ethics guidelines for trustworthy AI in 2019.
2. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, which developed a set of principles for ethical AI in 2016.
3. The Montreal Declaration for Responsible AI, which was developed by a group of AI researchers and practitioners in 2018.
4. The Asilomar AI Principles, which were developed by a group of AI researchers and practitioners in 2017.
5. NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) is intended to mitigate risks while designing, developing, using, and evaluating AI products, services, and systems.
6. Deon adds an ethics checklist to data science projects.



Digital Public Goods Alliance

unicef
for every child