

Tracing labour, power, and information in Artificial Intelligence Systems

Petter Ericson, AI Policy Lab, Department of Computing Science, Umeå University
pettter@cs.umu.se

1. Introduction

It is common for technology to be used to obscure the role of humans, and artificial intelligence (AI) is a field where this is even more true than for many others. From ghost workers and data scraping to algorithmic management and automated decision making, AI technologies are used to displace, appropriate, and hide human labour in various ways. Decisions are hidden inside technical systems, externalising them from individuals and organisations who could be held meaningfully accountable, which can make agency and power flow in new and often poorly understood ways. By having automated systems imitate humans, teleworkers can be seamlessly swapped in and out as needed with users none the wiser, making the systems appear significantly more capable than they actually are.

A useful abstraction for studying and exposing the workings of these systems is to consider how and where information is flowing through them. However, tools for such analyses tend to either be highly abstracted and focused on the broader sociotechnical systems where AI components are situated, or highly technical and focused on the details of specific software and hardware architectures, or on idealized and abstract models thereof. The proposed work will attempt to bridge this gap. On the one hand, it will take a more rigorous approach grounded in information theory to qualify and quantify the information that the humans involved share both through some specific technical system and through outside channels. On the other hand, it will take a wider view on the concrete workings of those technical systems, incorporating sociotechnical metadata into analyses of digital information flows.

In sum, we aim to use tools and methods from information theory and sociotechnical system modelling, together with formal graph models and complexity theory, to investigate and explain how diverse human labour and knowledge is discretized, datafied, and expressed throughout the development and deployment of different types and architectures of AI systems.

A complementary aim of this work is to build on and further develop existing research connecting information, computation, labour, and value, and how these concepts interact, specifically in the context of AI systems, making concrete contributions to interdisciplinary studies on AI and data work. A second major aim is to yield new insights into how to quantify and qualify information flows, through a focused study of sociotechnical systems involving AI components, where information and its flows in the digital realm can be directly studied, and comparisons can be made to both models and empirical realities of the social realm surrounding them.

Ultimately, we aim to investigate the following research questions: * RQ1 What types of information flows can be identified within different AI system architectures, and how can these be formally categorized? * RQ2 How do human actors contribute to and engage with information flows in and around AI systems, and how can these social interactions be systematically modeled? * RQ3 How can we develop and validate models of information flows in AI-based sociotechnical systems that integrate both technical and human components? * RQ4 How do modeled information flows reflect or reinforce particular organizational or institutional power structures?

2. Related work

Though to the best of the authors knowledge there is very little research on precisely the present topic, its interdisciplinary nature means that there are a number of intersecting areas of active research. In particular, works that cover the intersection between AI and information theory, between information theory and labour, between labour and AI, and between any of these three areas and sociotechnical modelling are all relevant.

For the first intersection, Jeon and Roy have recently investigated the connections between Bayesian machine learning and Shannon information theory, drawing an equivalence between the cumulative errors during a learning process of an optimal machine learning algorithm, and the amount of information contained in the data. From a different angle, several works such as Tseng et al. [2] and [3] have looked specifically at large language models (LLMs), drawing on compression and entropy calculations to study the use, and training, respectively of LLMs as related to natural language texts.

For the second, Dantas [4] has drawn direct connections between information and both labor and value in an explicitly Marxist framework, distinguishing between not only use and

exchange value, but also semiotic value, and deriving a specific notion of information work which will be of direct use in the proposed work. Dantas also draws a distinction between random and redundant information work, which is similar to the distinction between semantic and syntactic information work in [5] which is further nuanced into an explicit spectrum in [6].

The third intersection is itself a broad area, with many different aspects of relevance. Nguyen and Mateescu [7] gives a good overview of the current landscape in relation to Generative AI specifically, while Davis [8] provides a broader review of relevant issues, making an explicit (and useful) distinction between cases where AI use impacts labour demand (through automation) and those relating more to worker power (through surveillance, algorithmic management, and the like). Further, works such as Crawford [9], Miceli and Posada [10], Gray and Suri [11], Merchant [12], Sadowski [13], and Mejias and Couldry [14] are all relevant for further developing this work. The Data Workers Inquiry (<https://data-workers.org>) will be another important source of alternate perspectives on AI and labour.

In terms of studying AI sociotechnical systems, once again several active areas are of interest. In particular Wu et al. [15] has developed a framework for integrating various types of models of sociotechnical systems (STS) into a single meta-model. Several modelling languages for sociotechnical systems exist, including STS-ml [16], which was developed for cybersecurity applications, and the host of standards and notations related to Business Process Model and Notation (BPMN), such as Decision Model and Notation (DMN), which is particularly relevant for models integrating AI decision support systems. However, all of these abstractions and models tend to integrate assumptions that are not always helpful for the purposes of this work. A relevant example of how existing modelling framework scan be extended to cover new areas is [17], which adds properties and functionality to STS-ml in order to check sociotechnical systems for compliance with the EU General Data Protection Regulation (GDPR).

A relevant parallel effort, though not directly related to the work we propose here, is that of Gutierrez Lopez and Halford [18], who aim towards an extension of XAI principles that including the sociotechnical environment of the machine learning system.

3. Aims

The main contribution of this work will be to integrate previous work on sociotechnical systems modelling with several notions of information and labour, specifically in the context of artificial intelligence. An additional benefit of this work will be to lay a basis for a further analysis of agency and accountability: By studying the information flows and potential inputs and decisions from humans involved in an AI sociotechnical system, together with an analysis of the power relations among them, accountability and responsibility can be transparently and meaningfully assigned.

We hope to make meaningful contributions to the practical use of information theory and information flow, as well as yield actionable and concrete directions for further exploration of new AI sociotechnical architectures. As part of this, a major component will consist of qualitatively and quantitatively analysing the information flows into and out of AI and ML systems, which will also give new and useful insights into the design of Hybrid AI systems in particular.

By creating concrete tools and methods for tracing information flows through both technical and social layers of AI systems, this work will attempt to offer not just theoretical insight, but practical value for those developing, regulating, or critically analyzing such systems. In a time when the societal consequences of AI are increasingly opaque yet consequential, this research will provide actionable models that can inform transparency standards, system audits, and future AI governance efforts.

4. Preliminary results

4.1. Categories of information

A foundational topic of this work includes clarifying and classifying different types of information. In particular, though at extreme small ranges reality can occasionally appear to be digital, for most practical purposes, it is continuous. In contrast, digital information, and computer and AI systems, while implemented on physical hardware, are conceptually and practically discrete. As such, while abstractions of the human and physical sections and relations of a sociotechnical system are going to inevitably be lossy, for the digital parts it is

in principle possible to be both precise and concrete. This then, must be our first distinction between fundamentally different types of information: Abstracted notions and models of the real world, and concrete digital bits and bytes.

In terms of different theoretical notions of information, we further contrast the more mathematical definitions of Shannon [19] ('minimal code') and Kolmogorov [20] ('minimal program') with the Batesonian concept of 'a difference which makes a difference' [21]. A fourth relevant concept is Corning's 'control information' [22], which rather than connecting Shannon entropy/negentropy directly to the physical thermodynamic concepts with the same name, instead quantifies the amount of information contained in some signal or phenomenon by the amount of physical changes that it can effect. An example taken directly from Corning [22] is that of a car approaching a stoplight. If the driver does not notice or understand the traffic light, there is no control information being transferred by whatever light is shown. Only if the driver both sees the light, understands it, and is prepared to change the future trajectory of the car, is there any control information being sent out by the light switching to red. Broadly, we can thus consider two very different types of information flows: The almost entirely discrete and abstract digital information exchanges between and inside of software components, and the messy, socially situated, and necessarily contingent and abstracted information flows that can be modelled to exist between humans, technological artefacts, and their surrounding physical context. The main interest of this work lies precisely where these information flows intersect and interact.

4.2. Human-computer information interactions

With a minimal distinction of information flows as above, consider the interactions between a human and a computer system: the transfer of information from human to computer will necessarily abstract some concrete intention of the human into a concrete digital signal, but likewise a (digital) computer output will take on a specific meaning to the human which depends on their prior knowledge and the context in which the output is given. We can depict these shifts as in figure 1.

4.3. Analysis example

As an illustrative example of the type of analysis that we aim to make more concrete, detailed, and empirically grounded, consider the case of an article being written about a sports event. It is, at this point, plausible that such an article could be written by a large language model (LLM) given an appropriate prompt, including some sort of summary of the 'relevant facts' of the event in question (e.g. the final tally of points, who made them when, and any injuries and other specific incidents, which are accessible from some sort of API). The situation would look something like figure 2. We can complicate this picture, however, by adding more context. The article will not reach publication without an editor, and the hidden labour that has resulted in the LLM is entirely absent in our initial figure, as is the work to set up the "sports API" and the later work to feed it with the 'relevant facts' from observations of the event. A more realistic picture emerges, as in figure 3.

Compare this to a situation where a human writer is the author of the same article. Though the plain (abstract) facts of the event in question may be the same, the human will also have access to an infinitely larger context as part of their writing process, both through direct experience and memory, and through communications with other humans, computer systems, and physical objects such as books and recordings, nevermind the sights, sounds, and smells of the event itself if the writer was also present at the event. In this case, the situation will look more like figure 4. This too can be made more complex, particularly if we imagine the writer to make use of an LLM for writing assistance of some sort, yielding a situation as in figure 5.

5. The path forward

Though primarily based in computing science, the nature of the problems addressed by the work call for an interdisciplinary approach. Notably, by building on existing work in Science and Technology Studies, as well critical marxist literature, it is possible to better situate and analyse the information flows and AI sociotechnical systems inside existing societal power structures and socioeconomic realities. In addition to the various theories of information mentioned in Section 4.1, we will also distinguish between different types of information work, as outlined in Section 2. The distinctions between data, information, and knowledge have been explored e.g. in [23], [24], and these perspectives will also be considered.

We will primarily be building on existing frameworks for the analysis of program structure and information flow through software. Notably, the theory and practice of Quantitative Information Flow (QIF) analysis in computer security, though focused on detecting and plugging information leaks between public and private variables under static source code analysis conditions, offers a range of useful tools for modelling intentional information flows as well. From a software engineering lens, constructing program flow graphs and clearly and consistently delineating components in a software system is a well established practice, with a host of frameworks available for use. An example of an abstract framework for describing program and information flows that has been specifically developed for purposes of describing Hybrid AI systems is the boxology of Harmelen and Teije [25].

For situating a software system in an organisational context, tools from business modelling are available as well, with well-established frameworks such as BPMN, STS-ml and various derivatives having seen extensive use to analyse information flows and decision processes in business contexts. Concretely, the near parts of this work will consist a phase of conceptual and theoretical grounding, studying and comparing existing frameworks for information flow analysis, to arrive at a rigorous and flexible framework for modelling information flows in sociotechnical systems, incorporating the above distinctions and specificities, and giving specific attention to questions of agency and valuation. This work will aim at identifying connections and distinctions in how different frameworks frame decisions, and how labour is considered within them.

In the course of this development, a metadata schema for information and information flows will be developed that can describe and categorise information both in terms of its qualities, its different information contents, as well as its role at a specific point in a described sociotechnical process. Tracing the changes of these properties as attached to a particular piece of information will be an important complement to the analyses of the flows themselves, and of the various transformations imposed on and driven by the information.

During and after these developments, the framework and schema will be empirically applied to real-world cases, both existing ones from the literature, and new and comparable studies of previously understudied sociotechnical contexts. Modelling these flows will be accomplished through direct study of technical artefacts and their documentation, as well as organisational policies and descriptions of their surrounding sociotechnical contexts. These will be supplemented by interviews and surveys of involved stakeholders to elicit new and undocumented perspectives not previously represented even in internal documents.

Through comparative analysis across multiple cases (to be selected to reflect diversity in AI architecture and deployment and across different domains, e.g. public-sector automation, language models, decision-support tools) the framework will be further refined to capture how different system configurations mediate flows of information, labour, and power across different AI configurations. Ultimately, we aim for a formal, extensible modelling framework for analyzing information flows in sociotechnical systems involving AI, as well as a richly annotated library of concrete case studies. Additionally, we aim to make both conceptual and methodological contributions to the study of accountability, power, and labour in AI, as well as help drive further developments in related fields.

References

- [1] H. J. Jeon, B. V. Roy, Information-Theoretic Foundations for Machine Learning, 2024. URL: <http://arxiv.org/abs/2407.12288>. doi:10.48550/arXiv.2407.12288, arXiv:2407.12288 [stat].
- [2] Y.-H. Tseng, P.-E. Chen, D.-C. Lian, S.-K. Hsieh, The semantic relations in LLMs: An informationtheoretic compression approach, in: T. Dong, E. Hinrichs, Z. Han, K. Liu, Y. Song, Y. Cao, C. F. Hempelmann, R. Sifa (Eds.), Proceedings of the Workshop: Bridging Neurons and Symbols for Natural Language Processing and Knowledge Graphs Reasoning (NeusymBridge) @ LRECCOLING-2024, ELRA and ICCL, Torino, Italia, 2024, pp. 8–21. URL: <https://aclanthology.org/2024.neusymbridge-1.2/>.
- [3] M. Yin, C. Wu, Y. Wang, H. Wang, W. Guo, Y. Wang, Y. Liu, R. Tang, D. Lian, E. Chen, EntropyLaw: The Story Behind Data Compression and LLM Performance, 2024. URL: <http://arxiv.org/abs/2407.06645>. doi:10.48550/arXiv.2407.06645, arXiv:2407.06645 [cs].
- [4] M. Dantas, Information as Work and as Value, tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society 15 (2017) 816–847. URL: <https://www.triple-c.at/index.php/tripleC/article/view/885>. doi:10.31269/triplec.v15i2.885.
- [5] J. Warner, Labor in information systems, Annual Review of Information Science and Technology 39 (2005) 551–573. URL: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/aris.1440390120>. doi:10.1002/aris.1440390120.

- [6] J. Warner, The spectrum of semantic and syntactic labour, *Journal of Documentation* 80 (2024) 649–664. URL: <https://www.emerald.com/insight/content/doi/10.1108/JD-03-2023-0057/full/html.doi:10.1108/JD-03-2023-0057>.
- [7] A. Nguyen, A. Mateescu, *Generative AI and Labor: Power, Hype, and Value at Work*, Technical Report, Data & Society Research Institute, 2024. URL: <https://datasociety.net/library/generative-ai-and-labor>. doi:10.69985/gksj7804.
- [8] O. F. Davis, *Artificial Intelligence and Worker Power* (2024).
- [9] K. Crawford, *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*, Yale University Press, New Haven London, 2021.
- [10] M. Miceli, J. Posada, The Data-Production Dispositif, *Proceedings of the ACM on Human-Computer Interaction* 6 (2022) 1–37. Publisher: ACM New York, NY, USA.
- [11] M. L. Gray, S. Suri, *Ghost work: how to stop Silicon Valley from building a new global underclass*, Houghton Mifflin Harcourt, Boston, 2019.
- [12] B. Merchant, *Blood in the machine: the origins of the rebellion against big tech*, first edition ed., Little, Brown and Company, New York, 2023. OCLC: on1389775757.
- [13] J. Sadowski, *The mechanic and the luddite: a ruthless criticism of technology and capitalism*, University of California Press, Oakland, California, 2025. doi:10.1525/9780520398085.
- [14] U. A. Mejias, N. Couldry, *Data grab: the new colonialism of big tech and how to fight back*, WHAllen, London, 2024.
- [15] P. P.-Y. Wu, C. Fookes, J. Pitchforth, K. Mengersen, A framework for model integration and holistic modelling of socio-technical systems, *Decision Support Systems* 71 (2015) 14–27. URL: <https://www.sciencedirect.com/science/article/pii/S016792361500007X>. doi:10.1016/j.dss.2015.01.006.
- [16] E. Paja, F. Dalpiaz, P. Giorgini, Modelling and reasoning about security requirements in socio-technical systems, *Data & Knowledge Engineering* 98 (2015) 123–143. URL: <https://www.sciencedirect.com/science/article/pii/S0169023X1500052X>. doi:10.1016/j.datak.2015.07.007.
- [17] C. Negri-Ribalta, R. Noel, N. Herbaut, O. Pastor, C. Salinesi, Socio-Technical Modelling for GDPR Principles: an Extension for the STS-ml, in: *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, 2022, pp. 238–243. URL:

<https://ieeexplore.ieee.org/document/9920163/?arnumber=9920163>.

doi:10.1109/REW56159.2022.00052, iSSN: 2770-6834.

[18] M. Gutierrez Lopez, S. Halford, Explaining machine learning practice: findings from an engaged science and technology studies project, *Information, Communication & Society* 28 (2025) 616–632. URL:

<https://www.tandfonline.com/doi/full/10.1080/1369118X.2024.2400130>.

doi:10.1080/1369118X.2024.2400130.

[19] C. E. Shannon, A mathematical theory of communication, *The Bell System Technical Journal* 27(1948) 379–423. URL: <https://ieeexplore.ieee.org/document/6773024>.

doi:10.1002/j.1538-7305.1948.tb01338.x, conference Name: The Bell System Technical Journal.

[20] A. N. Kolmogorov, On Tables of Random Numbers, *Sankhyā: The Indian Journal of Statistics, Series A* (1961-2002) 25 (1963) 369–376. URL:

<http://www.jstor.org/stable/25049284>, publisher: Springer.

[21] G. Bateson, Form, substance and difference, *Essential readings in biosemiotics* 501 (1970). Publisher: Springer.

[22] P. A. Corning, Control information theory: the ‘missing link’ in the science of cybernetics, *Systems Research and Behavioral Science* 24 (2007) 297–311.

URL:<https://onlinelibrary.wiley.com/doi/abs/10.1002/sres.808>. doi:10.1002/sres.808, _eprint:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/sres.808>.

[23] L. Businska, I. Supulniece, M. Kirikova, On Data, Information, and Knowledge Representation in Business Process Models, in: R. Pooley, J. Coady, C. Schneider, H. Linger, C. Barry, M. Lang (Eds.), *Information Systems Development*, Springer, New York, NY, 2013, pp. 613–627. doi:10.1007/978-1-4614-4951-5_49.

[24] L. Businska, I. Supulniece, Towards Systematic Reflection of Data, Information, and Knowledge, *Scientific Journal of Riga Technical University. Computer Sciences* 43 (2011). URL: <https://content.sciendo.com/doi/10.2478/v10143-011-0002-9>. doi:10.2478/v10143-011-0002-9.

[25] F. v. Harmelen, A. t. Teije, A Boxology of Design Patterns for Hybrid Learning and Reasoning Systems, *Journal of Web Engineering* 18 (2019) 97–124. URL:

<http://arxiv.org/abs/1905.12389>.doi:10.13052/jwe1540-9589.18133, arXiv:1905.12389 [cs].

Date: June 24, 2025

Source: <https://aipolicylab.se/aipex/>

Keywords (comma separated):

information theory, information flow, socio-technical system modelling

Related URL (if any):

https://people.cs.umu.se/~pettter/tracing_information_figures.pdf