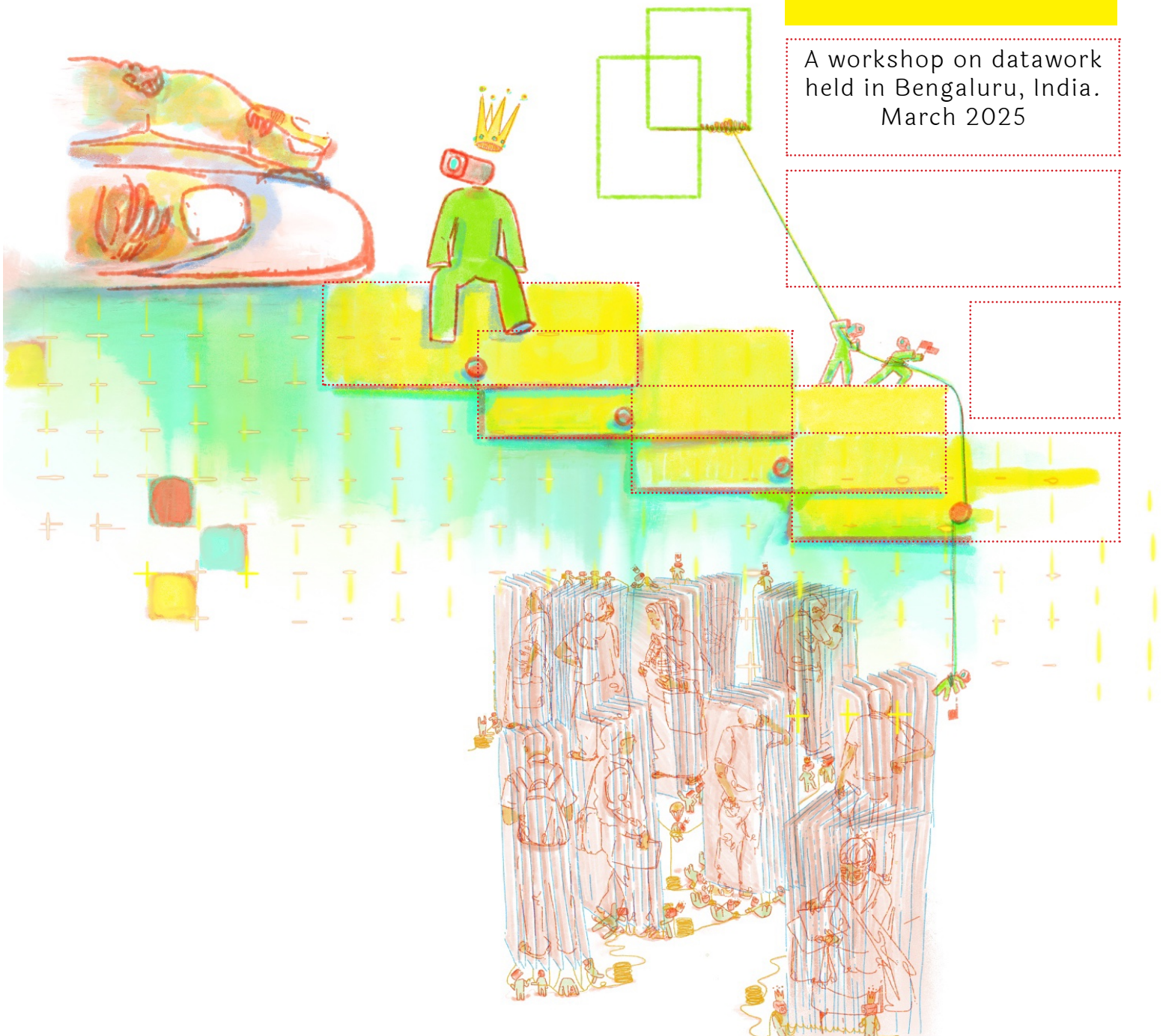


mapping Data Work

A workshop on datawork
held in Bengaluru, India.
March 2025



Mapping Data Work
A workshop on data work

First Edition, 2026

2026 Chandhiramowuli, Ausloos, Carraro, Singh. Licensed under CC BY 4.0.

This publication is licensed under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).
Under the terms of this license, users may copy, distribute, transmit, adapt,
and build upon the material for any purpose, including commercial use,
provided appropriate attribution is given to the original author(s) and source,
a link to the license is provided, and any modifications are clearly indicated.

License details: <https://creativecommons.org/licenses/by/4.0/>

Workshop organized by:

[Jef Ausloos, Srravya Chandhiramowuli, Aditya Singh,
Valentina Carraro, Priyam Vadaliya, Ritvik Gupta,
Tarunima Prabhakar, Kaustubha Kalidindi]

Workshop Research & Report by:

[Srravya Chandhiramowuli, Jef Ausloos,
Valentina Carraro, Aditya Singh]

Design, Layout & illustrations
[paulie]

Typography

-Titles: Solway Ex B

Designed by Mariya Pigoulevskaya (The Northern Block).
Licensed under the SIL Open Font License 1.1 (OFL).

-Body: Dosis

Designed by Pablo Impallari, Edgar Tolentino, and contributors.
Licensed under the SIL Open Font License 1.1 (OFL).

-alt Body: Laila

Designed by Hitesh Malaviya and Jonny Pinhorn
for Indian Type Foundry.
Licensed under the SIL Open Font License 1.1 (OFL).

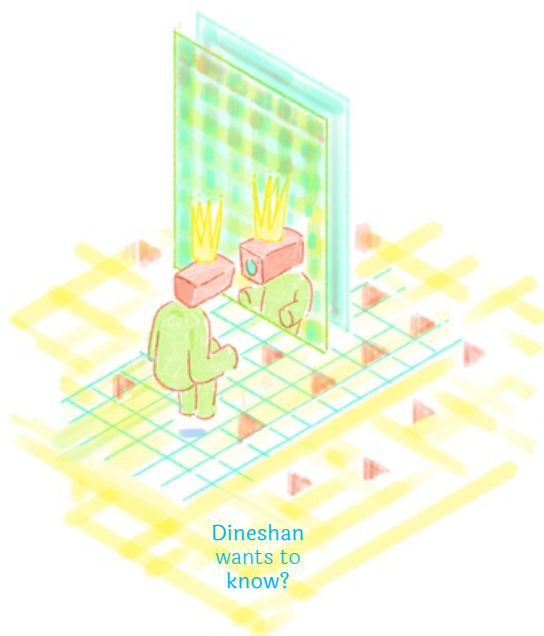
Publication Details

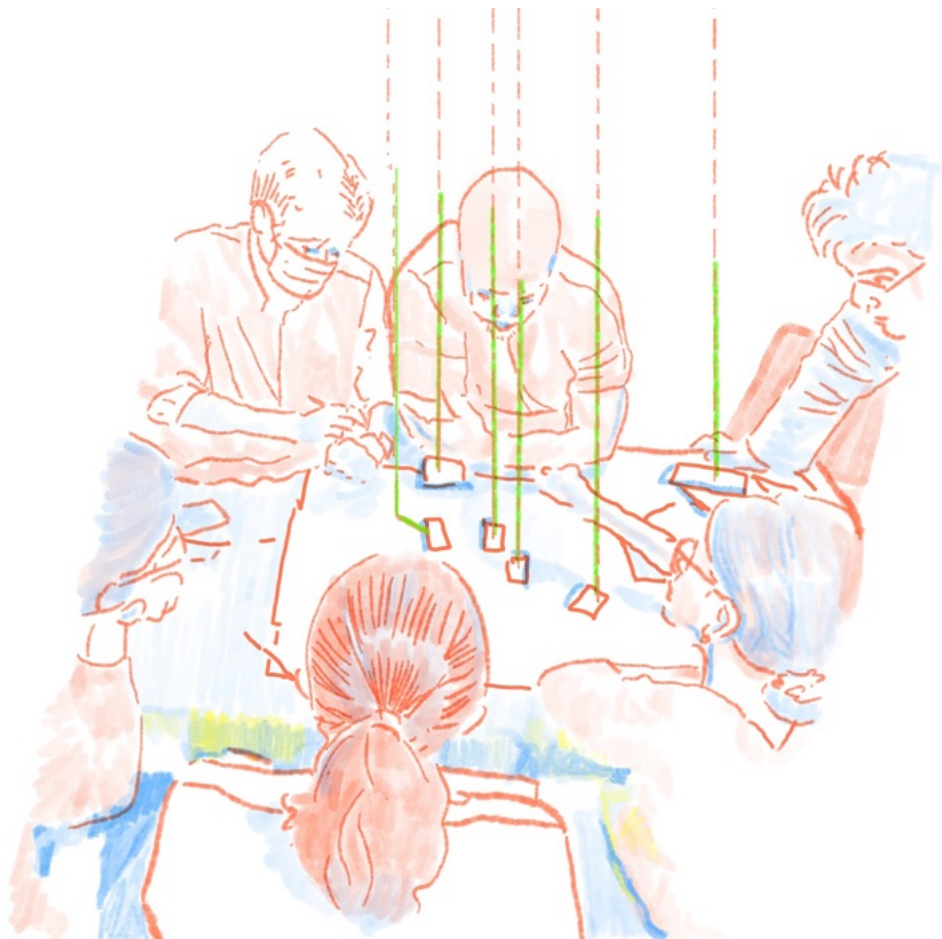
This publication is part of a *veni*-project financed by the Dutch Research Council (NWO) under the grant C.2127.0303.

Website:
<https://mappingdata.work>

mapping Data Work

A workshop on data work
held in Bengaluru, India.
March 2025





FOREWORD

This report documents the proceedings of Mapping Data Work, a workshop held in Bengaluru in March 2025, bringing together diverse perspectives to examine the human labour that underpins AI systems but remains largely invisible in dominant narratives of technological progress. Bengaluru, as both a major site of outsourced data work and a hub of AI ambition, offered a fitting location to explore these tensions.

The workshop was convened in a spirit of open-ended inquiry, drawing on research, lived experience, and critical practice to ask:

- Who builds AI,
- Under what conditions, and
- With what consequences
for workers, knowledge systems,
and global justice?

Contents

1. Introduction	pg 9
2. Workers' concerns: Precarity and devaluation of data work	pg 15
3. Automation, outsourcing, and surveillance in data work	pg 19
4. Human cogs in the AI machine	pg 23
5. From outsourced digital labour to data work	pg 29
6. Mapping accountability and Emerging sites of data work	pg 33
7. The tech sector response: Resisting regulation, offloading risks	pg 37
8. Frictions in valuing data work	pg 41
9. Alternative models of data and AI production	pg 45
10. Concluding Reflections and Next Steps	pg 49
Appendix	

WORD NET
IMAGE NET
FEI-FEI LI

2008-2010

AMAZON
MECHANICAL
TURK

160M - IMAGES

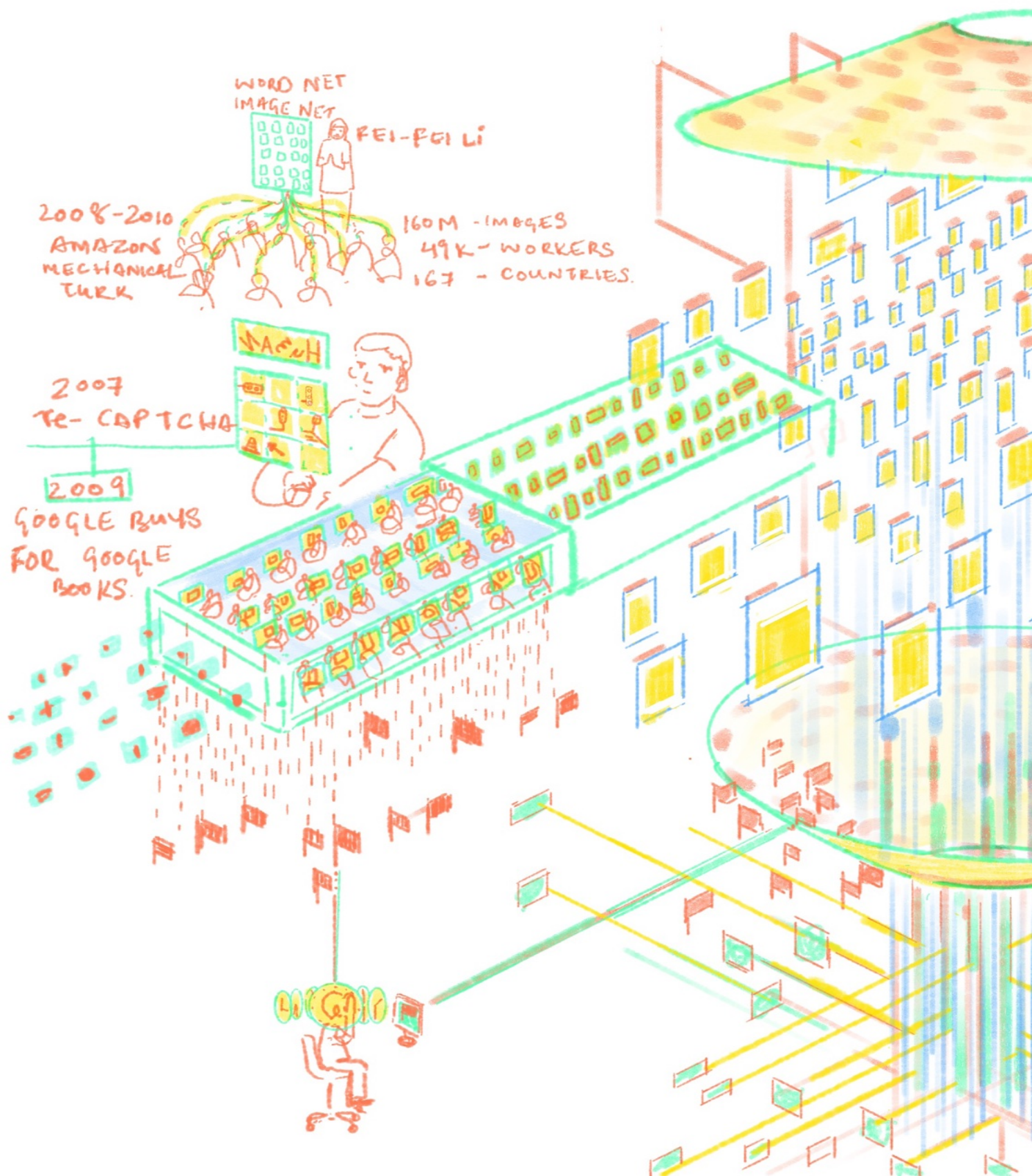
49K - WORKERS

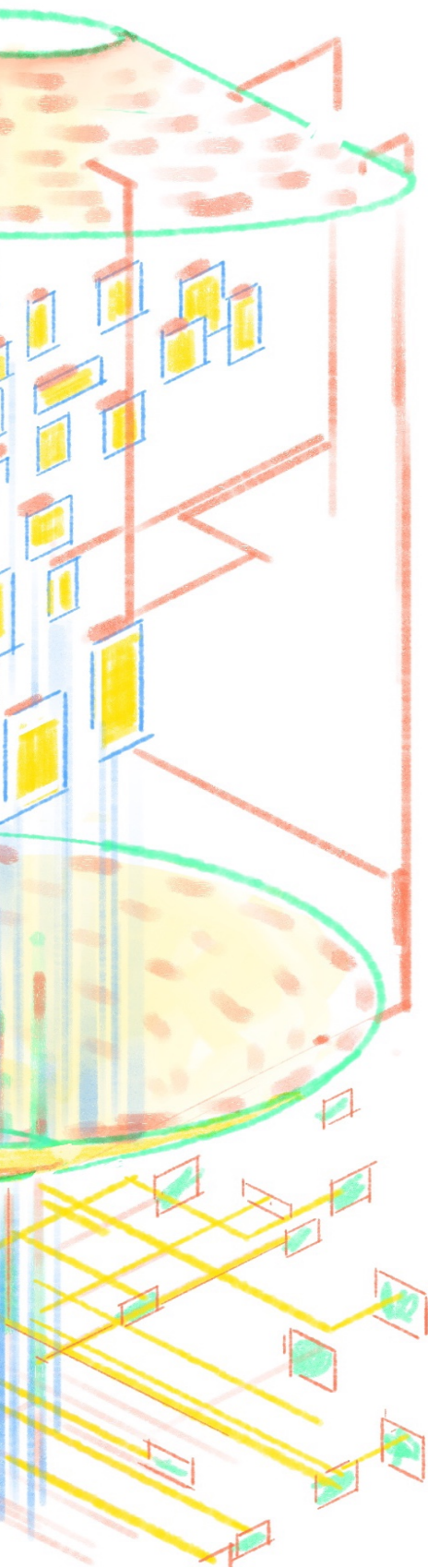
167 - COUNTRIES.

2007
TE-CAPTCHA

2009

GOOGLE BUYS
FOR GOOGLE
BOOKS.





Ch 1

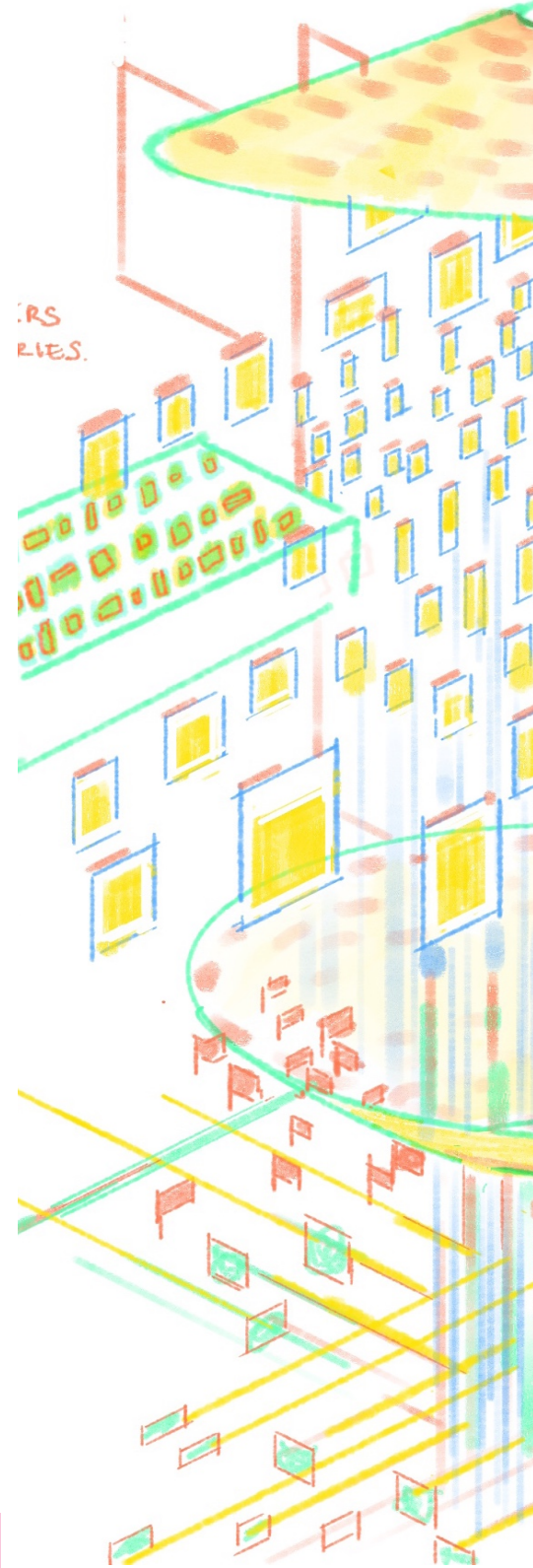
9

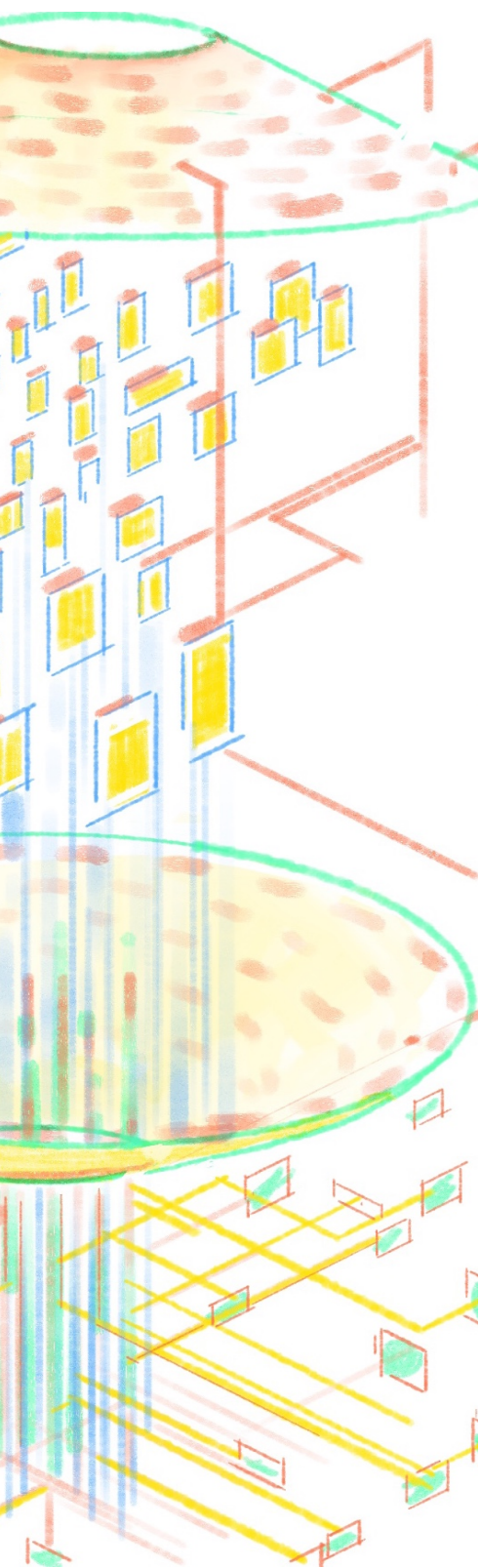
Introduction

Popular narratives around AI often focus on its potential to surpass human intelligence and transform professional domains and social contexts. AI development has received heightened attention, entrepreneurial interest and investment over the last few years, especially since the launch of generative AI models. Across different parts of the world, tech companies and nation states have rushed to embrace AI as necessary, inevitable, and desirable, for economic and social development. As efforts to scale, adapt and replicate AI models soar, fuelled by hype around AI, the crucial role of human input to train, fine-tune, evaluate, and moderate these systems remains largely obscured.

While state regulators and policymakers have been slow to recognise and account for the material dimensions of AI value chains, researchers and civil society organisations have long raised concerns about environmental destruction, exploitative labour conditions in raw material extraction, and the extractive logics of generative AI affecting creatives worldwide. These critical perspectives remain massively overshadowed by the well-funded promise of AI as a universal solution to contemporary societal challenges from climate change to poverty. Here, AI is constructed as producing an objective, almost ideal form of knowledge, while the skilled and careful labour of data workers that underpins the entire enterprise is rendered invisible — with the risk that AI systems end up reproducing and amplifying the dominant ontologies and taxonomies imposed by large corporations in the Global North.

ip: Ai and made magic : The beginnings of what has become ‘data labour’ and the evolution of an entangled system where the human and human’ness across contexts is unevenly mapped with the future promise of universal convenience.





Against this backdrop, in March 2025 we – researchers and practitioners at Universities of Amsterdam (NL), Edinburgh (UK), Aapti Institute (IN), and Tattle Civic Tech (IN), engaged with data work, data governance, and critical data studies – co-organised a workshop in Bengaluru, India. Together, our interest was in crafting the workshop as a space for an open-ended, critical exploration of data work with empirical insights, practice-based reflections, and perspectives from academia, policy, advocacy, design, and critical technical practice¹. This collaborative effort aimed to foreground the contributions and concerns arising from this oft overlooked aspect of AI production and examine the colonial legacies, epistemic concerns and labour implications of data work in India and broader the Global South.

Bangalore was a particularly fitting location to explore these themes. India is a major site of outsourced data work for the global tech sector. At the same time, its government and industry also share an ambition to position themselves at the forefront of digital transformations through investments in AI and digital public infrastructure (DPI). Bangalore's position as a tech hub and its active civil society and research ecosystem allowed us to situate the workshop within the productive tensions between the on-the-ground reality of data work and the aspirational narratives surrounding AI.

¹ The full list of workshop participants and organisers is included at the end of the report.

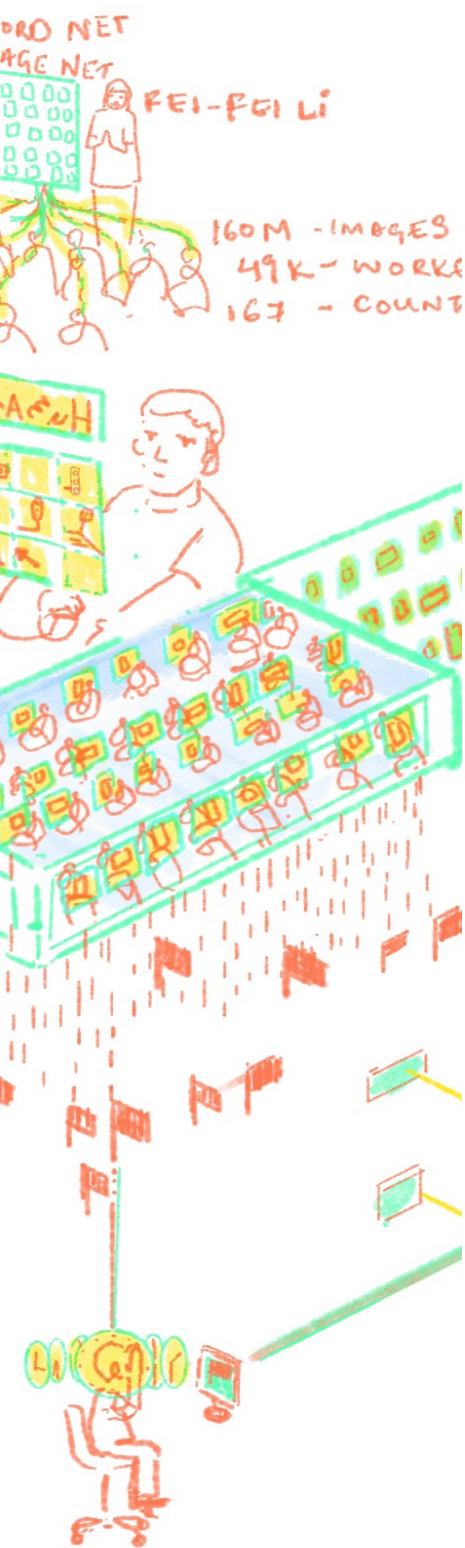
This workshop was designed to move beyond critique and techno- and euro-centric policy approaches. To do so, we brought together a very diverse group of participants from different continents, sectors and socio-economic backgrounds. Through presentations, discussions, interactive and creative exercises, we sought to stimulate a more open-ended exploration of what data work is, the issues around it, and how we might imagine positive paths forward.

This report attempts to document and summarise the collective wisdom of this gathering. **Key themes** of discussion at the workshop included:

- i. The exploitative and dehumanising conditions of labour in data work including unfair wages, delayed payments, precarious contracts and the lack of workplace protections, surfaced through the lived experiences of data workers as well as research into different modes and sites of data work.
- ii. Epistemic concerns regarding whose knowledge is codified or appropriated through AI systems, particularly in areas like agri-tech where knowledge of the farming community is discounted or delegitimised through the imposition of data- and AI-based systems.



p: People behind Pixels : There is a nascency to the labour of cleaning up and setting up a dataset through processes like tagging, annotation and content moderation for machine learning with different purposes. What started as BPOs are now multiple platforms and forms of data labour extraction, where consent and coercion are often hindsight reconfigurations, the material impact often real-time and felt individually as an affect of labour obscured behind the gloss of AI.

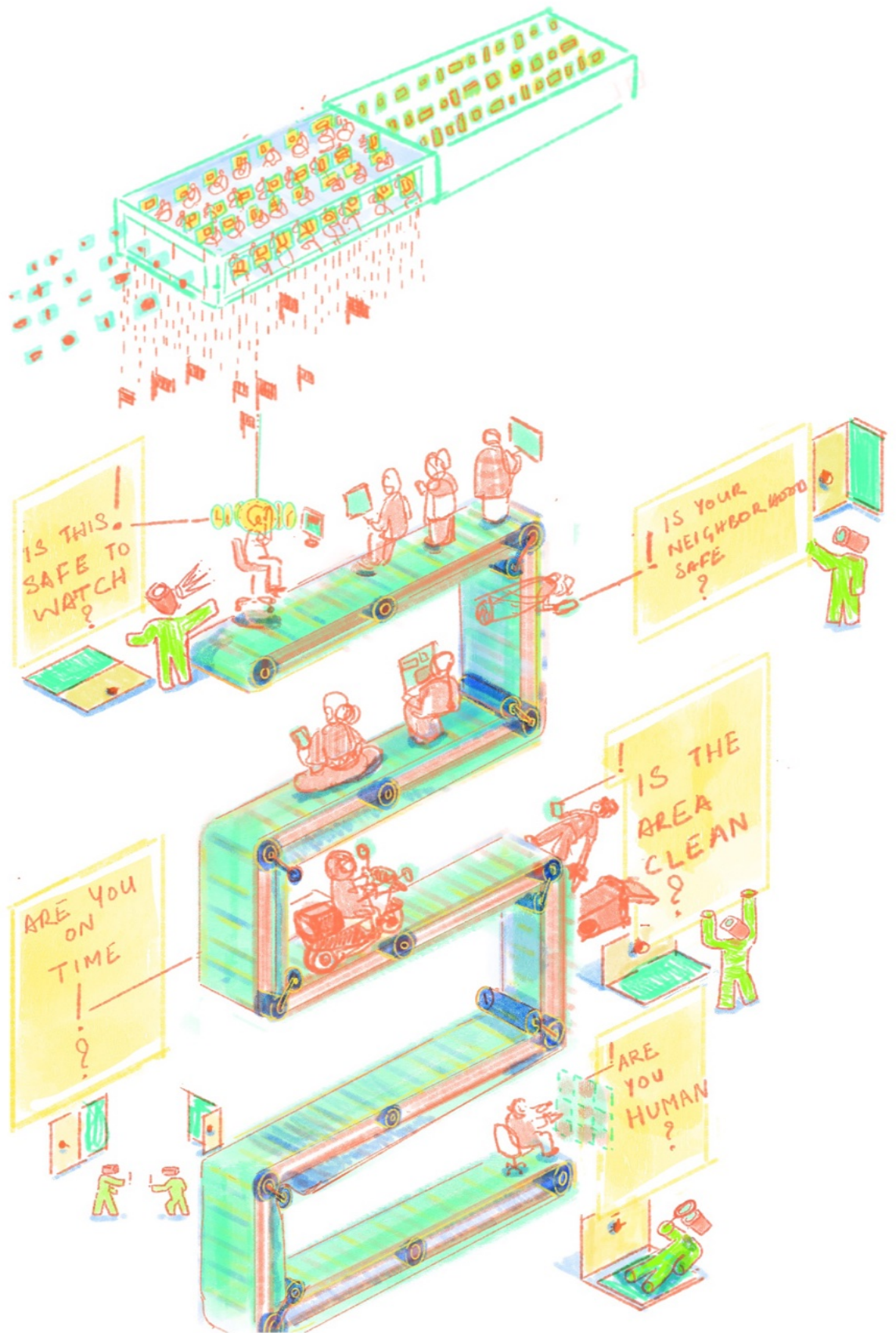


iii. A conceptual expansion of data work that accounts for the diverse, often invisibilised forms of labour imposed on a wide range of people—disabled users, government employees, and citizens accessing public services—as they navigate bureaucratic systems increasingly mediated by AI.

iv. This labour is structured through CAPTCHA tasks, algorithmic monitoring, and workplace surveillance tools, all in the service of maintaining AI systems adopted as markers of organisational innovation.

v. These varied sites of data work and the associated flows of data, capital, and labour begin to reveal the complex, trans-national supply chains that underpin data and AI production, inviting further attention to how they are structured and the challenges they pose for accountability.

Finally, alternative approaches to data and AI production that resist the exploitative tendencies in the sector, too, are caught within these supply chains, which present challenges to building ethical interventions within an extractive industry. Non-profit and social enterprise initiatives must confront the tensions in collaborating with or competing against more powerful, better-funded actors, and the risks they pose for limiting or diluting agendas for worker agency and ethical labour practice.



Workers' concerns: Precarity and devaluation of data work

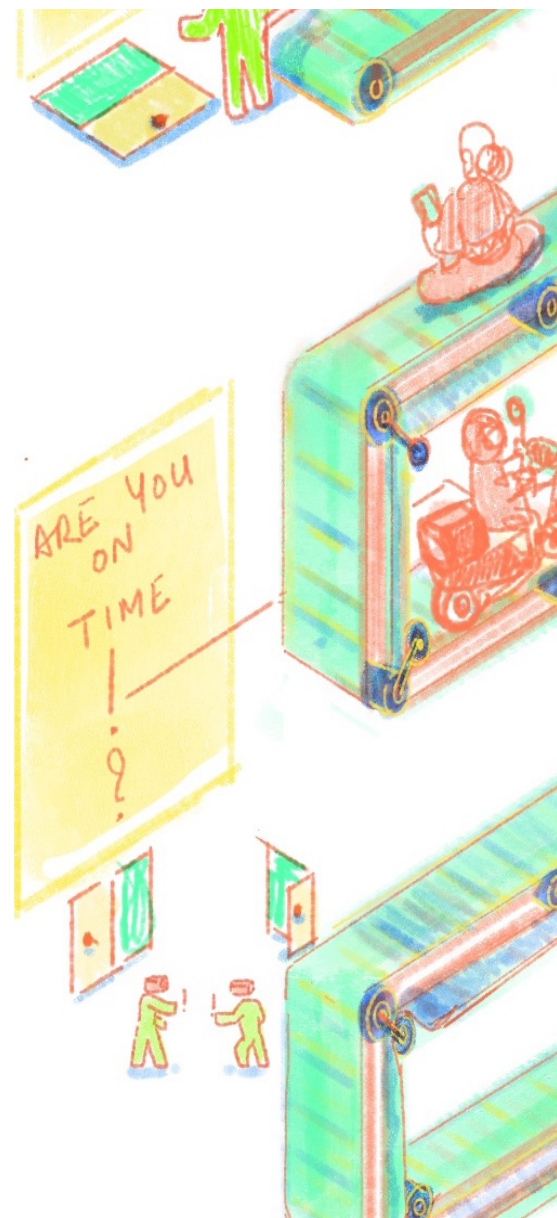
Ch 2

15

The workshop sought to centre the voices and lived experiences of data workers, and learn from the first hand accounts of navigating data work. We were joined by Mophat Okinyi (former data worker and union organizer, founder of Techworker Community Africa) and Jeevitha SatheeshKumar (senior data worker and entrepreneur, founder of JSR Script), who drew on their experiences of working in content moderation and data work to describe a fragmented and opaque ecosystem shaped by short-term contracts, unclear payment structures, and limited workplace protections.

Data workers often face delayed or denied payments following opaque quality checks, while layered vendor arrangements—where clients contract intermediaries who further subcontract the work—obscure the relationship between what workers are paid and the overall value of the contract.

This lack of transparency weakens bargaining power and contributes to downward pressure on wages, particularly in highly competitive markets where new entrants may accept lower rates out of necessity.

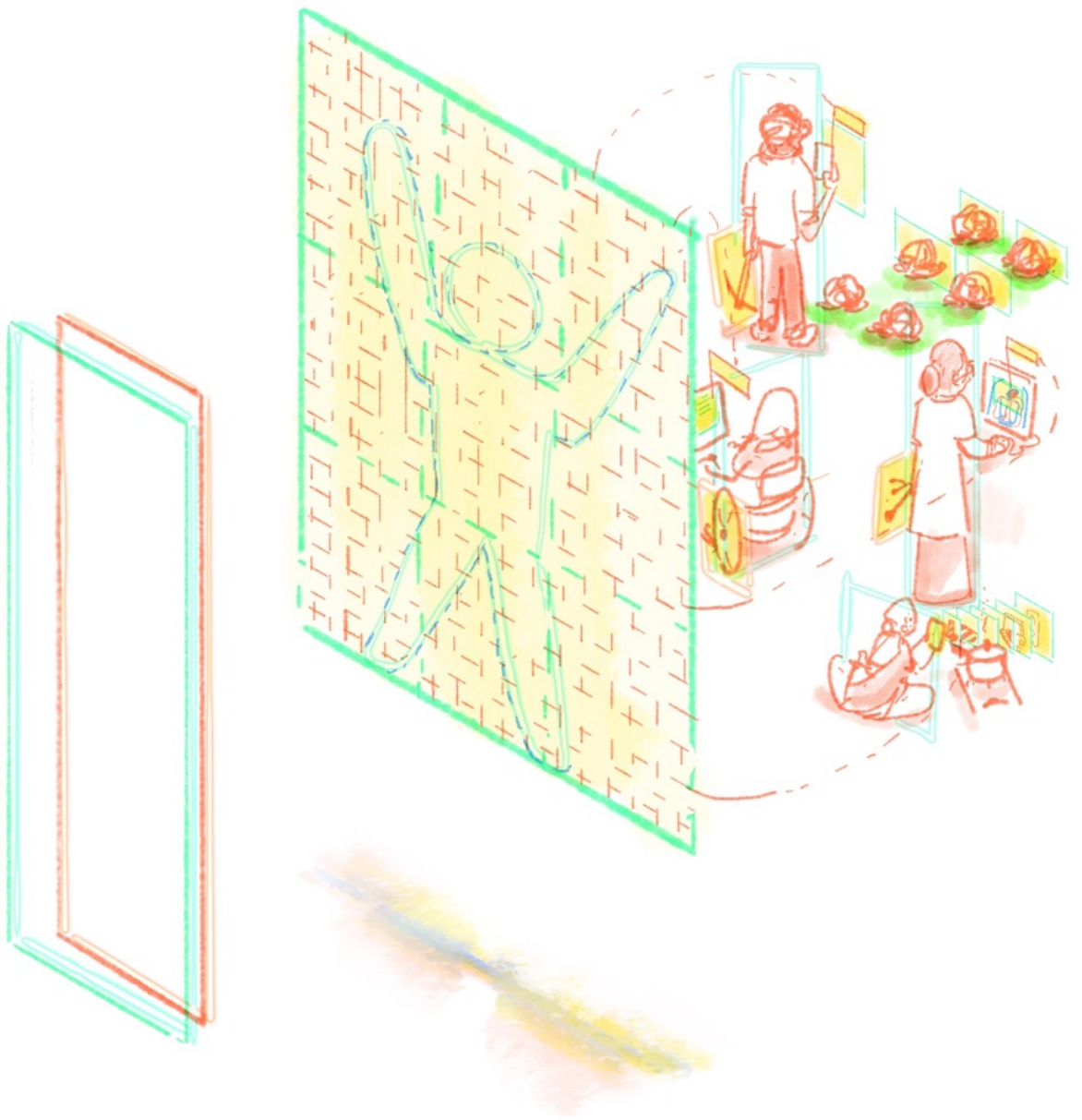




p: Convenience-Clickwork and the Conveyor of Bad News: A lot of clickwork that frames the emergence of AI often obscures the mundane yet labouring work-flows behind curating datasets, guiding machine vision and other 'world-model' creation for the AI layer across multiple contexts. The conveyor belt of clickwork-outliers is an attempt to indicate how precarious these jobs are often letting go of and harvesting behavioral data of workers at their worst moments, as benchmarks of what and who is newly framed as 'productive and product'. Everything else is mapping risk or Bad News for business as data.

The panelists also challenged the persistent framing of data work as low-skilled “clickwork,” emphasising the concentration, contextual judgement, and domain knowledge required for tasks such as annotation, transcription, translation, and conversational data generation. Content moderation in particular carries a heavy psychological toll, with workers navigating exposure to disturbing material under strict performance metrics that often exceed nominal working hours.

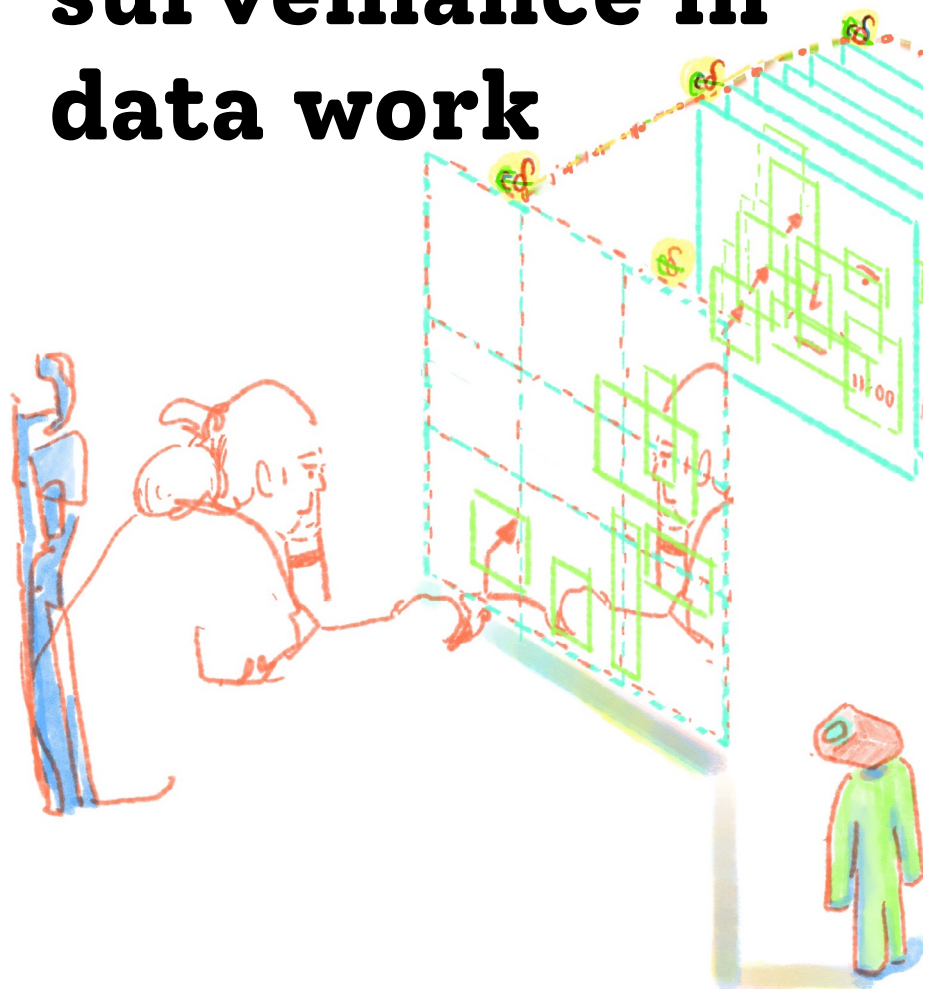
Gendered vulnerabilities further compound these challenges, as women workers tend to face harassment and heightened precarity in short contract cycles. While worker-led initiatives and associations are advocating for fairer pricing, clearer contracts, mental health support, and stronger labour protections, organising remains difficult amid a geographically dispersed and oft-isolated workforce, employer resistance and limited government enforcement—especially where political and economic interests are closely tied to the expansion of digital labour markets.



Automation, outsourcing, and surveillance in data work

Ch 3

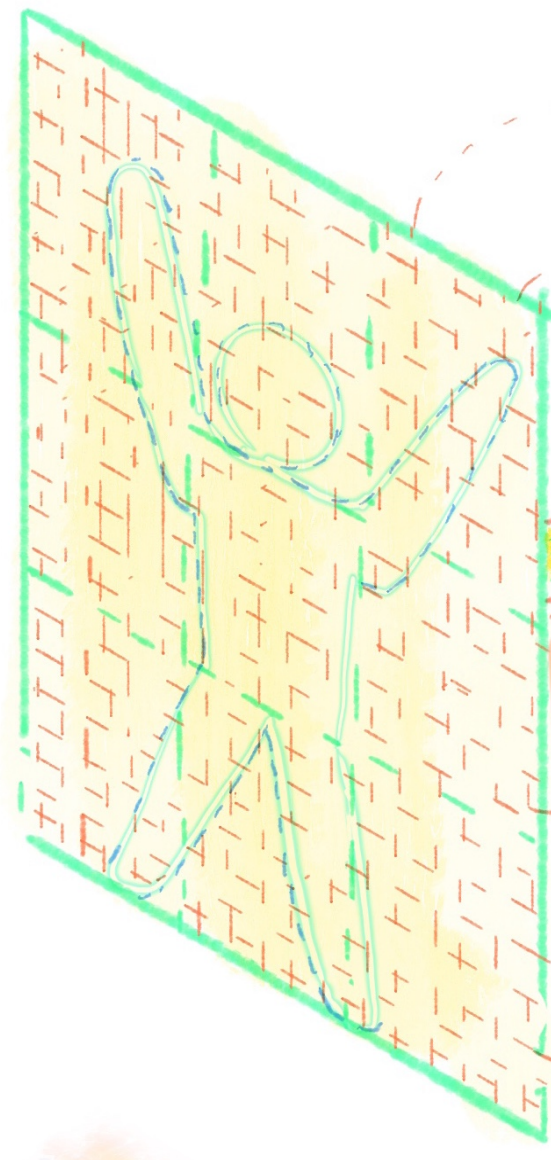
19



Some of the previous themes were further explored and extended by presenters examining the impacts of data work through design, policy, advocacy and research. Research by Meera Muthukrishnan (IIIT Bangalore), Swati Ganeshan (IIIT Bangalore), and Joyeeta Dey (Manipal University) examined CAPTCHA tasks as exclusionary data labour.

The presentation reframed CAPTCHA as a form of unpaid data work, showing how systems designed to verify “humanness” both extract value for companies and impose disproportionate, often inaccessible burdens, especially on disabled users. What appears as a simple security step is in fact exploitative and exclusionary labour, raising urgent questions about accessibility, fairness, and who truly benefits from everyday human-machine interactions.

Riya Ravishankar and Shreya Toshniwal’s (Fold Labs) presentation zoomed in on the possibilities for design interventions to foster worker agency and control in data work. They probed how data design toolkits could be implemented in data work interfaces and integrated into the design process itself, invoking discussions about the role of design in interrupting prevailing data work practices.



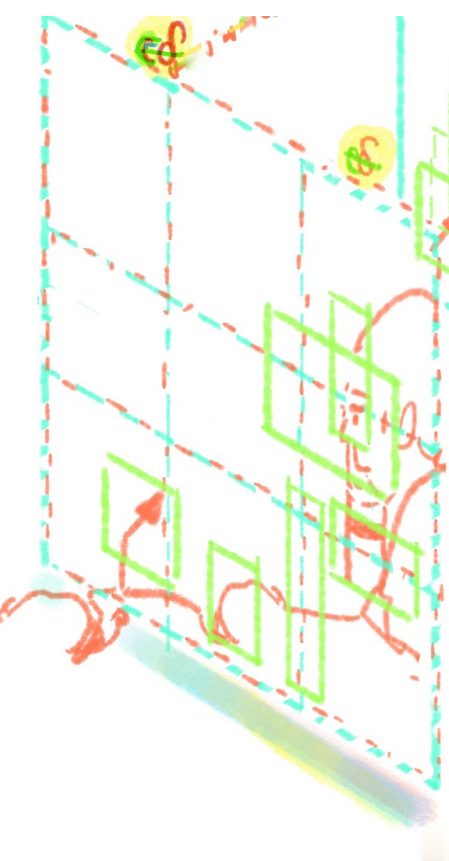
p: Completely Automated Public Task by Corporations for free Human Activity and mapping what dineshan wants to know : CAPTCHA work may have started as a way of proving human’ness in its initial configuration. For machine vision now, both the tasks and the new benchmarks of what ‘human’ness is, ‘productive work’, ideal ‘user’ of the service are also what gets framed, calculated, incentivised, designed and indexed for the future, from user to citizen. The CCTV camera-headed Dineshan is a corporeal stand in for surveillance-capitalism, another manifestation of digital technology and the citizen as data subject, as the nation-state always wants to know and the machine always wants to learn, the illustrations have some dineshan’s with crowns and some without, Big and Gig Dineshan, as it were.

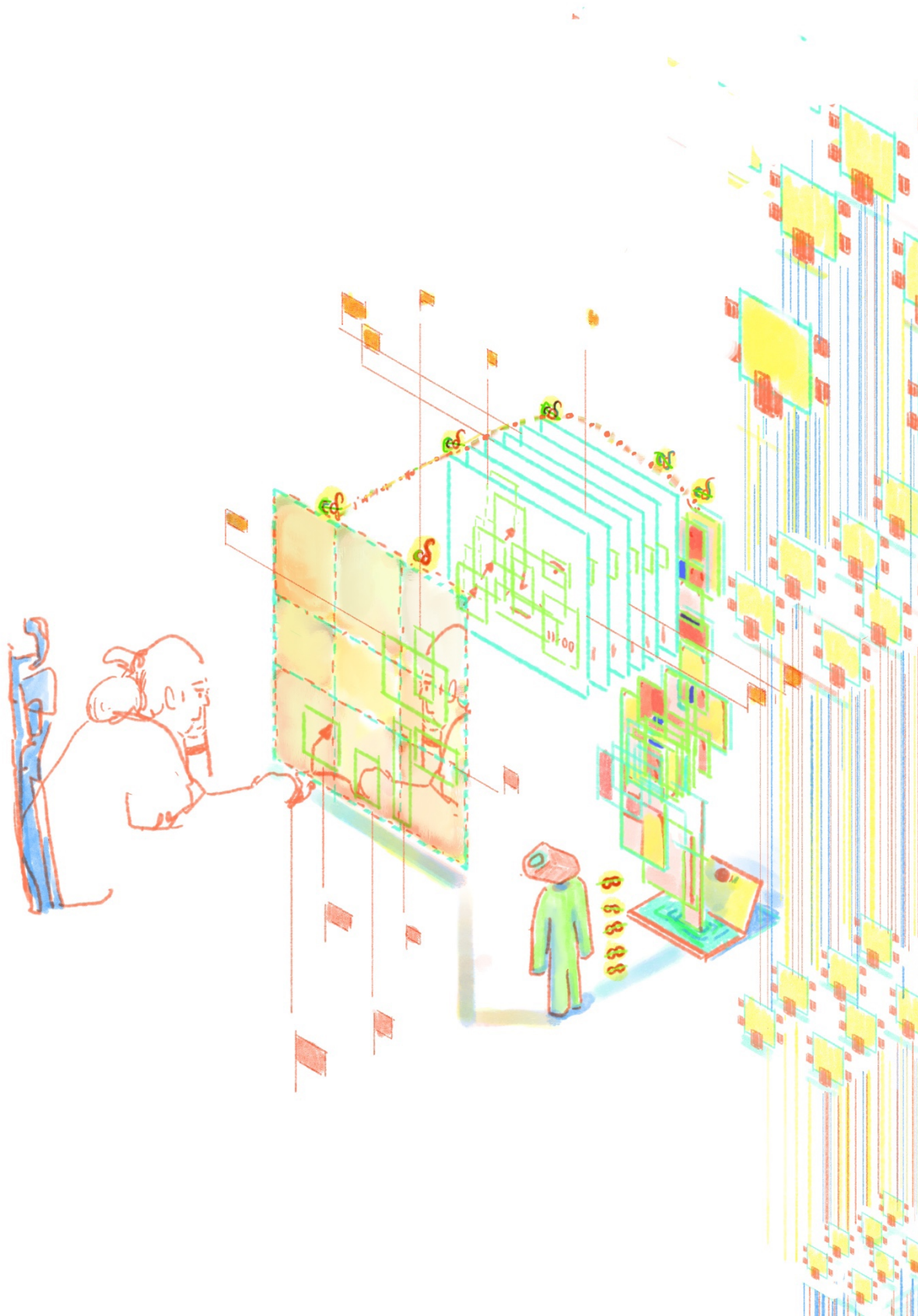


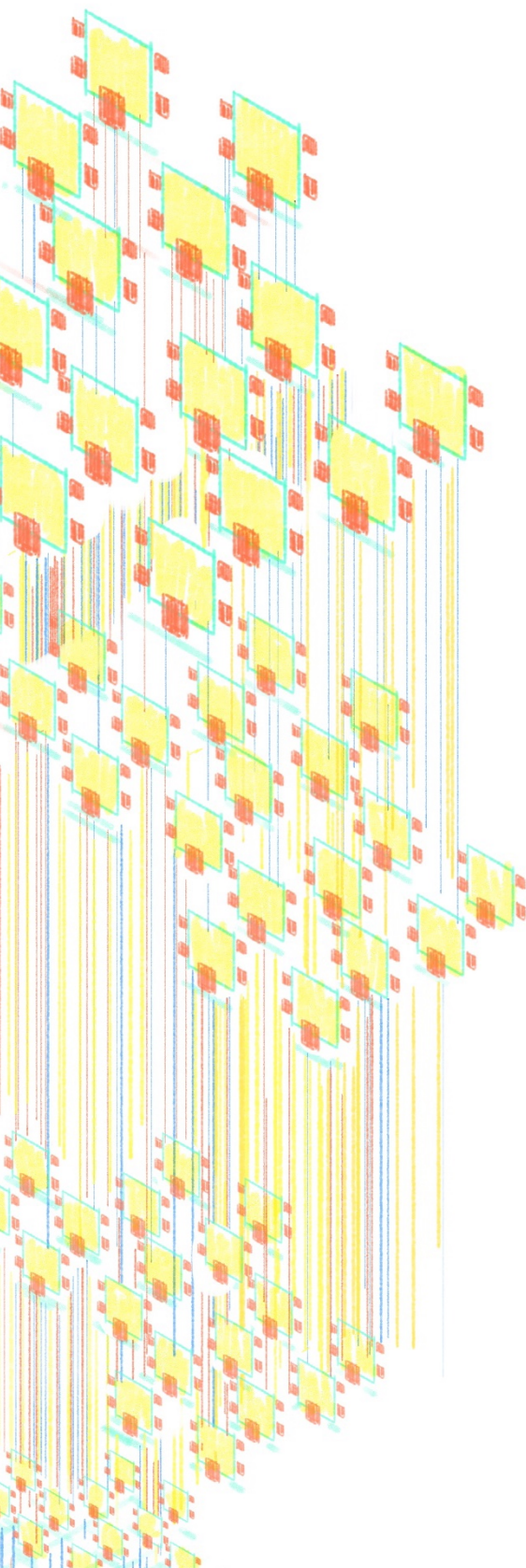
Abhineet Nayyar and Chetna VM (Centre for Internet and Society) presented the challenges in mapping the opaque ecosystem of AI data annotation firms that includes diverse organisational forms, clients, and loosely defined “AI” roles that complicate systematic analysis. They emphasised that understanding data work requires clarifying and mapping this complex organisational terrain, without which, assessing labour conditions or accountability remains fundamentally constrained.

The final presentation by Thomson Chakramakkil (IIT Delhi) examined the datafication of worker emotions through enterprise surveillance tools. He showed how systems like SCALA, an LLM-powered “employee engagement” chatbot, transform workplace monitoring into the extraction of affective labour, where employees’ emotions are captured, quantified, and repurposed as data to optimise control, segmentation, and profitability. Far from fostering genuine engagement, these tools instrumentalise trust and simulate empathy to convert human sentiment into managerial data capital—reshaping both surveillance and the role of management itself.

Together the papers productively complicate what constitutes data work, who undertakes the work, and the broader supply chain within which it operates. They also raised pertinent questions about how data work might be scoped, what its bounds might be as well as whether and how we might distinguish between different forms of data work as undertaken by users of digital platforms, content creators on social media, and workers at BPOs and platforms.







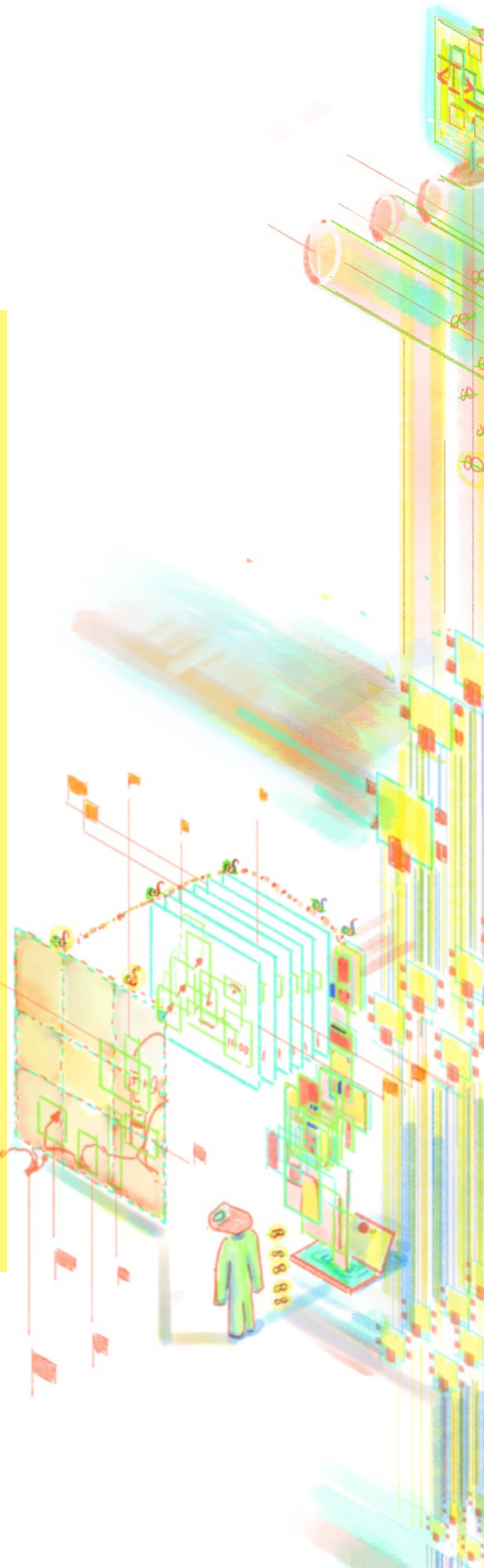
Human cogs in the AI machine

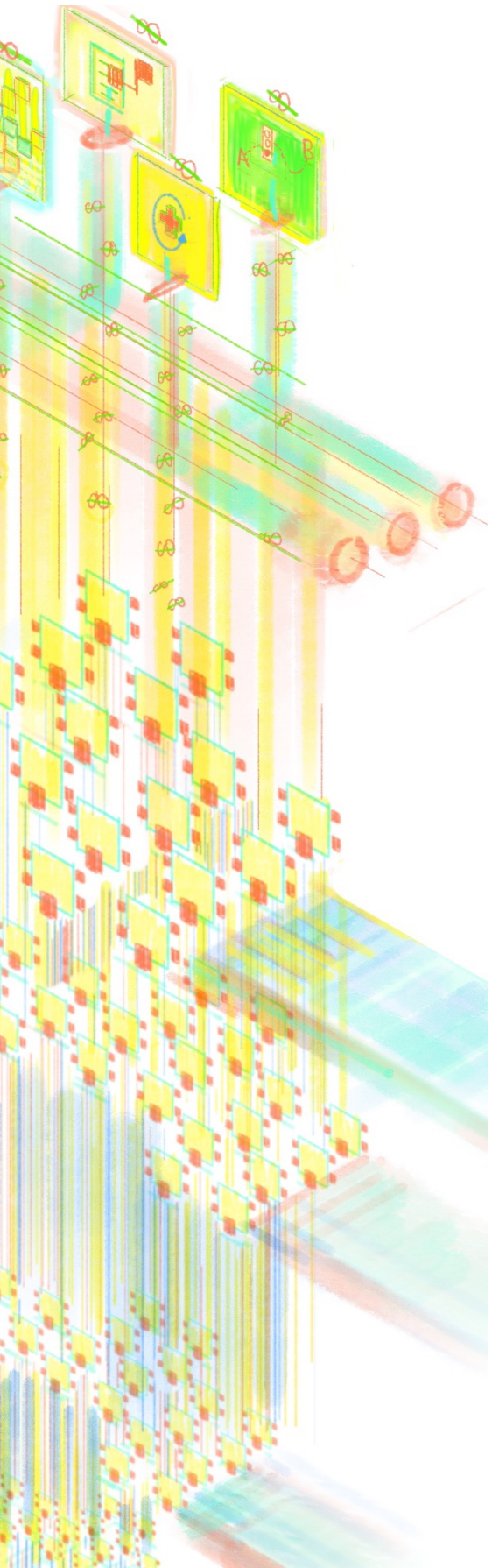
Ch 4

23

Building on the themes from the previous discussion, Dr. Uma Rani, Senior Economist at the International Labour Organisation (ILO) delivered a keynote lecture. The lecture drew attention to the relationship between algorithmic labour (developing, testing, and maintaining AI systems) and data labour (annotation, labelling, and moderation), arguing that although public discourse highlights the former, the latter is indispensable yet systematically invisibilised.

Drawing on a survey study conducted in Kenya and India, Dr. Uma Rani challenged narratives of AI-led “post-work” futures by documenting the human labour embedded across four AI domains: **large language models, autonomous vehicles, content moderation systems, and medical imaging tools**. Human input remains crucial at multiple stages, including creating demonstration datasets to fine-tune LLMs, remote operators intervening when autonomous vehicles encounter situations algorithms cannot resolve, and medical professionals producing ground-truth annotations for diagnostic AI. The keynote highlighted how rushed or undervalued data work leads to errors that propagate through systems, posing serious risks in areas like healthcare.



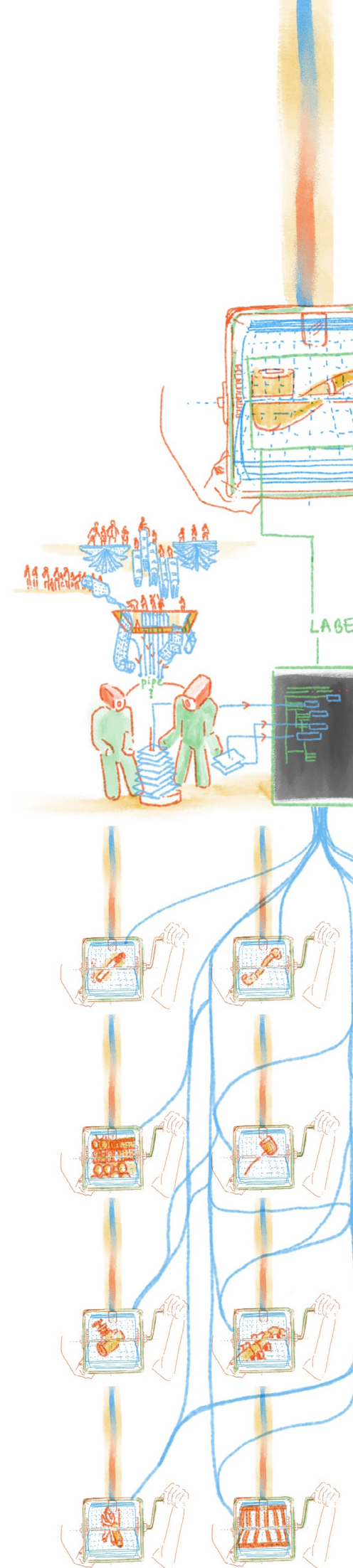


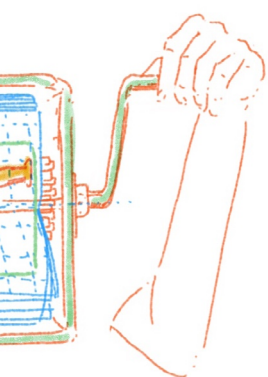
p: From surveillance to rentier capitalism of AI and disproportionate human labour : The services and conveniences layer with agentic models and realtime machine learning, highlights the emergence of a familiar layer of credits, tokens and recurrent running of dead-end loops without shifts in outcomes or impact, across distinct contexts, that affect and will continue to shape the future. The credits at the layer of technology and governance infrastructure is the real currency or data capital, marked with infinity, to show the scale of the gift. The credits or currency for everyone else, user and data worker alike is [dineshanal credit](#), marking those that use it or receive it as othered from the layer that promises [The Future](#), at the cost of the futures of everyone. Here the Yellow square with Red Flags depicts the data labourer simultaneously being mapped, person and the pixelation of human behaviour.

Their survey findings revealed that despite high educational attainment—nearly all Indian BPO data workers hold STEM bachelor's degrees and about half of Kenyan workers hold or are pursuing degrees—pay typically ranges from \$2–\$4 per hour across micro-platform and BPO jobs. Kenyan content moderators must make decisions in under a minute, often without adequate breaks, while NDAs prevent workers from discussing their tasks or psychological toll, including exposure to violent material.

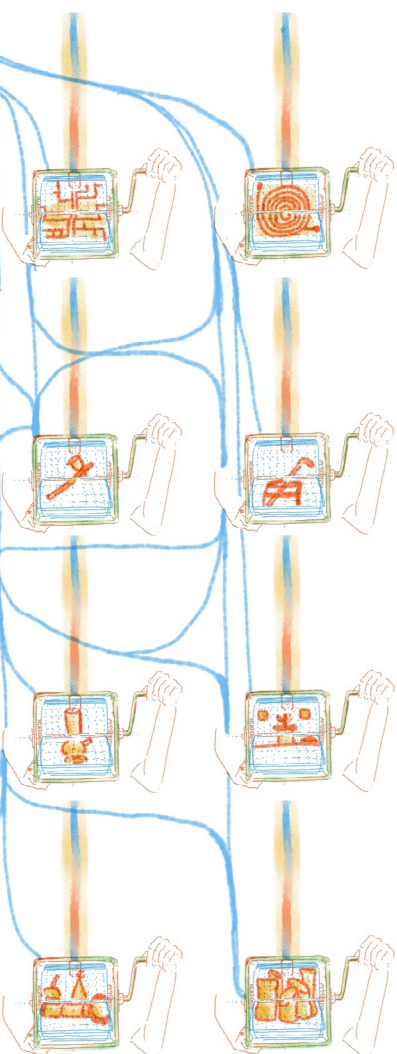
Recruitment practices are frequently misleading, with applicants expecting roles such as translation or data science instead end up performing annotation or moderation. Data work, thus, complicates development narratives that link education to mobility, as these jobs demand credentials but offer limit advancement, low pay, and poor conditions, prompting calls for stronger regulation of venture capital-driven AI development, expanded AI auditing, and employment policies that ensure digitalisation and AI production produces dignified work rather than “digital sweatshop” labour in the Global South.

The Q&A began by discussing a core contradiction in data work: although humans are integrated into AI pipelines to supply the nuance algorithms lack, workers themselves are constrained by rigid decision trees and classification protocols that require them to behave “like algorithms.” Dr. Rani noted this tension helps explain why areas such as content moderation cannot be fully automated, since human communication, cultural context, and local nuance exceed what models can reliably interpret — a limitation that also applies in high-stakes sectors including finance, healthcare, and law.





L: pipe, 42.0%



Participants further discussed how to mobilise the growing body of evidence on data labour, with Dr. Rani rejecting minimum wage as an appropriate benchmark given the high value attributed to data by the AI industry and the fact that many workers hold advanced technical degrees, and concluding with a call to shift from documenting harms to supporting unions and worker organisations in demanding better conditions while exploring alternative ownership models for AI that could transform the structure of data work.

p: Human learning Machine learning Human : Originally the illustration was commissioned*to describe machine vision in automated decision making systems (ADMS) used by the Indian Government. The visual metaphor of a mutoscope/hand cranked flip-book, is a referencing to how each frame of human activity is used to fill in the blanks of symbol grounding that machine vision still needs in terms of nuance. Behaving like algorithms and how each frame obscures and persists in terms of human vision and memory, while informing machine vision and new benchmarks of what is human.

*Ai Observatory by Divij Joshi: <https://ai-observatory.in>



From out- sourced digital labour to data work

Ch 5

29

Despite the rhetoric of disruption and novelty surrounding AI, much of what powers it is itself not new: transcription, data entry, verification, and content moderation have been performed by workers in the Global South for at least two decades.

To explore the continuities between older forms of out-sourced digital labour — such as call centre and back-office work — and what is today called “data work”, we convened a panel with Sona Mewati (IIHS), Aditya Ray (UWE Bristol) and James Oyange (African Content Moderators Union), on historicising data work.

A key theme of the discussion was that, while the terminology surrounding this work has evolved, labour conditions have remained largely unchanged: workers across BPOs, moderation firms, and crowdsourcing platforms continue to face wage disparities, opaque employment arrangements, and limited awareness of the broader systems their labour sustains.

The panelists challenged the “myth of AI” as autonomous, emphasising that human judgment—particularly in areas like content moderation—has historically been crucial to projects of digitisation and automation, and continues to be indispensable in the context of AI.





p: Tech Corporations as Nation States as Tech Corporations : The theories around the social contract evolved from the divine right of kings, today's social contract is slowly being reconfigured around the divinity that comes with the promise of AGI and Super Intelligence. memewhile human judgment will have to piece together what being human in the evolution of the contract might cost, as collectives and collective action is tested through new forms and means.



The discussion also foregrounded the gendered dimensions of data work, noting how women disproportionately undertake this labour, often from home, where its demands for care and emotional labour are undervalued even as it provides vital income. While efforts to organise remain difficult due to precarious employment, corporate resistance, and state narratives linking tech work to development, panelists pointed to emerging grassroots organising and peer networks among workers as early steps toward collective action.



Mapping accountability and Emerging Sites of Data Work

Ch 6

33



Mapping accountability in data work

The collaborative accountability mapping exercise, led and facilitated by Aapti, brought participants together in four groups to examine distinct segments of the data work ecosystem: content moderation and data labelling, each conducted either through BPOs or directly via digital labour platforms.

This structure enabled participants to explore shared accountability challenges while recognising the unique dynamics of each context.

The exercise followed three stages: first, mapping key stakeholders across the data work supply chain and considering how geography shapes their relationships; second, analysing decision-making processes, including who makes key choices and how factors such as legislation and labour markets influence them; and third, assessing how these decisions impact workers' everyday experiences, while identifying opportunities to improve fairness, wellbeing, and accountability across stakeholders and regions.

Each group then shared their findings in a brief shareout session that followed the activity.



p: Unevenly spread histories, futures, technologies and their recurring costs : What is emerging through conversations is the scale of variation in terms of how human labour is divided across the term 'artificial intelligence', the conversations are now also piecing together the division of the labourers themselves, based on old and new colonial geographies, as ever, shaping subsistence, justice and welfare in new systems of capital.

Emerging sites of data work

This session brought together case studies examining less-studied or overlooked forms of data work, expanding the frame beyond platform and BPO contexts that much of the previous discussions focused on.

Sanjana Malhotra (Columbia University) presented ethnographic research examining data work embedded in government bureaucracy. She highlighted how bureaucrats and contractors were tasked with photographing and geotagging infrastructure, producing data later used for AI models, often through ad-hoc labour arrangements. Amidst Indian ministries being evaluated for AI adoption, data work involved generating large volumes of low-quality data, turning the process into compliance-driven “data production for its own sake.”

Shruti Kaushik (India Kanoon) elaborated on the data work required to extract and classify legal citations from judgments, reaching about 83% accuracy after nine model revisions. She brought attention to how data work is never finished—even in structured domains like law—and open public datasets risk being scraped and absorbed into commercial AI systems such as ChatGPT without attribution or compensation.

Sara Marcucci (AI + Planetary Justice Alliance) highlighted the epistemic shifts that underpin AI in agriculture, comparing two datasets—a government crop recommendation dataset and a satellite imagery dataset developed in partnership with Google—against traditional farming knowledge, and finding that community-held insights on soil, pests, and seasonal practices are largely missing or poorly represented. This showed how rapid agricultural datafication risks erasing local knowledge and may turn “inclusion” into surveillance or commodification rather than benefiting farming communities.



The tech sector response: Resisting regulation, offloading risks

Ch 7

37

In a conversation between Sonam Jindal (Partnership on AI) and Valentina Carraro (University of Amsterdam), the discussion examined accountability in AI supply chains, reflecting on efforts both inside technology companies and in multi-stakeholder advocacy. Jindal noted that responsibility for data workers' conditions is highly diffused within tech firms: no single team typically owns the issue (or responsibility), and even sympathetic employees often lack the authority to implement changes.

Efforts to improve practices have focused on cultivating internal champions and developing data-sourcing guidelines through collaborative processes with major AI developers, but such voluntary frameworks face limits because they rarely reach key internal actors—including finance, legal, and product teams—whose decisions shape labour conditions.

It was also highlighted that this belies a US-centric view that approaches accountability as a negotiation with power-laden actors in a political environment where stronger regulation remains uncertain despite its potential effectiveness.

The discussion pointed out that substantive change in the data work sector requires shifting AI companies' risk calculations by making the costs of exploitative data practices more visible; one of the most persuasive levers is quality, since poorly paid, overworked data workers generate lower-quality training data that becomes expensive to correct later.





More fundamentally, the speakers observed that many AI systems are economically viable only because data labour is artificially undervalued, and they called on academia to move beyond documenting harms toward producing detailed, actionable research on supply chain structures, contract arrangements, and leverage points where interventions could more immediately influence company behaviour.

p: All that Glitters and All that is Gold : There is a dearth of policies, legal frameworks and good-faith actors when it comes to ensuring the mental and physical well-being of data workers, this is despite acknowledging how crucial the nuance they bring at each stage can be and should be- collective wisdom and not proprietary knowledge, if it helps prevent or manage burnout and worse. Information and knowledge metaphors around enlightenment and bringing to light that which was previously unknown, is also the basis for a lot of UX Copy behind AI.

Here the light is a distraction from the collective human labour behind automation and machine learning, and how a division of labour and labourers, individuates suffering under the framing of overall 'progress'.



Frictions in valuing data work

Ch 8

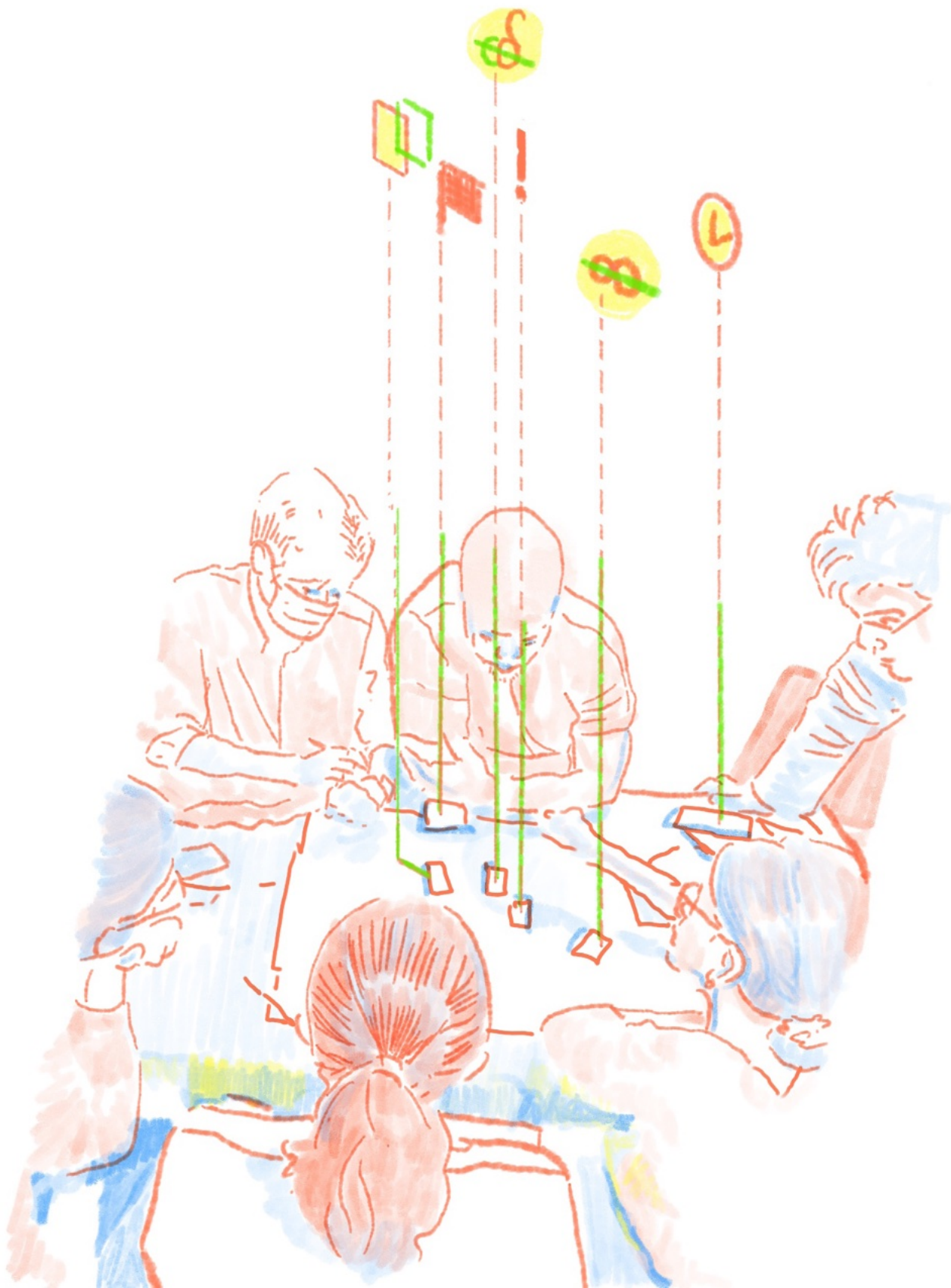
41

We explored the tensions in ethical data work practice in this group activity-based session, titled 'An Offer One Can't Refuse' facilitated by Tattle, that invited participants to collectively explore how data work is valued, compensated, and recognised across different contexts. The exercise prompted participants to simulate the complex decision-making involved in appropriately valuing and compensating data work.

Participants were assigned roles within a data annotation supply chain and asked to respond to hypothetical project offers under constraints of time, budget, and quality expectations. As the scenario evolved, teams negotiated trade-offs between fair compensation, speed, and dataset quality, revealing the structural pressures that shape real-world data labour practices and sparking discussions about how ethical principles might be adopted in actual practice.

The activity is made available to be played or adapted:
<https://uli.tattle.co.in/an-offer-one-cant-refuse/>

p: **Contexts and Main Quests** : The budgeting exercise really makes it clear how much downward pressure for data workers exists, when it comes to cutting corners around mental health, volume of work and underquoting to ensure projects don't get scalped. The pragmatics of it also shapes the expendable nature of data workers, with poor pay, little rest and sense of how click work is evolving with everyday use of machine learning themselves.





Alternative models of data and AI production

Ch 9

45

The final session of the workshop brought together practitioners building alternatives to extractive data work models. The panel featured Chandana Dhanraj from Karya, a social enterprise committed to ethical dataset creation, and Nihar Desai, from ArtPark's Project Vaani, an inclusive speech dataset initiative covering the linguistic diversity of India at the district level.

Karya strives to offer well above minimum wage, flexible work that does not disrupt workers' primary livelihoods, and in some cases implementing a data ownership model where workers receive royalties when datasets are reused. Ethical practice, here, also includes limits on screen time, attention to psychological safety, refusal of certain tasks such as content moderation, and long-term trust-building through partnerships with local bodies and panchayats; the organisation also conducts language documentation projects, including collecting women's spoken accounts of recipes and food knowledge in endangered languages.

Project Vaani described its district-level effort to build a representative speech dataset across India, based on the insight that linguistic variation does not align with the simplified language categories typically used in the tech industry, with significant differences occurring both between and within regions; achieving quality therefore requires intentional inclusion of underrepresented communities even when this complicates trade-offs between time, cost, and scale.

The discussion highlighted tensions in attempting ethical interventions within an extractive industry, including concerns that such initiatives may merely patch a flawed system or that non-profit and social enterprise models may be outcompeted by better-funded actors, as well as deeper questions about whether datafying marginalised languages

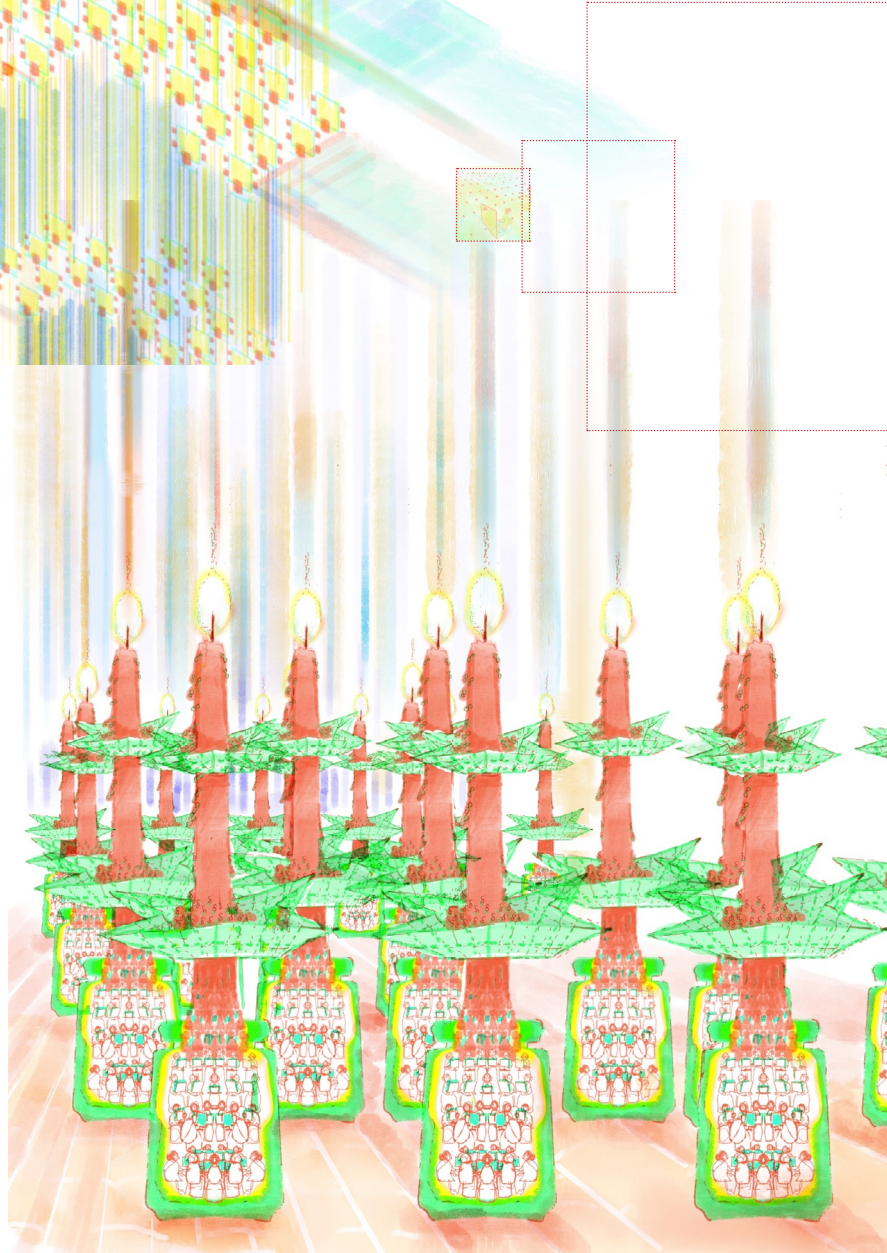
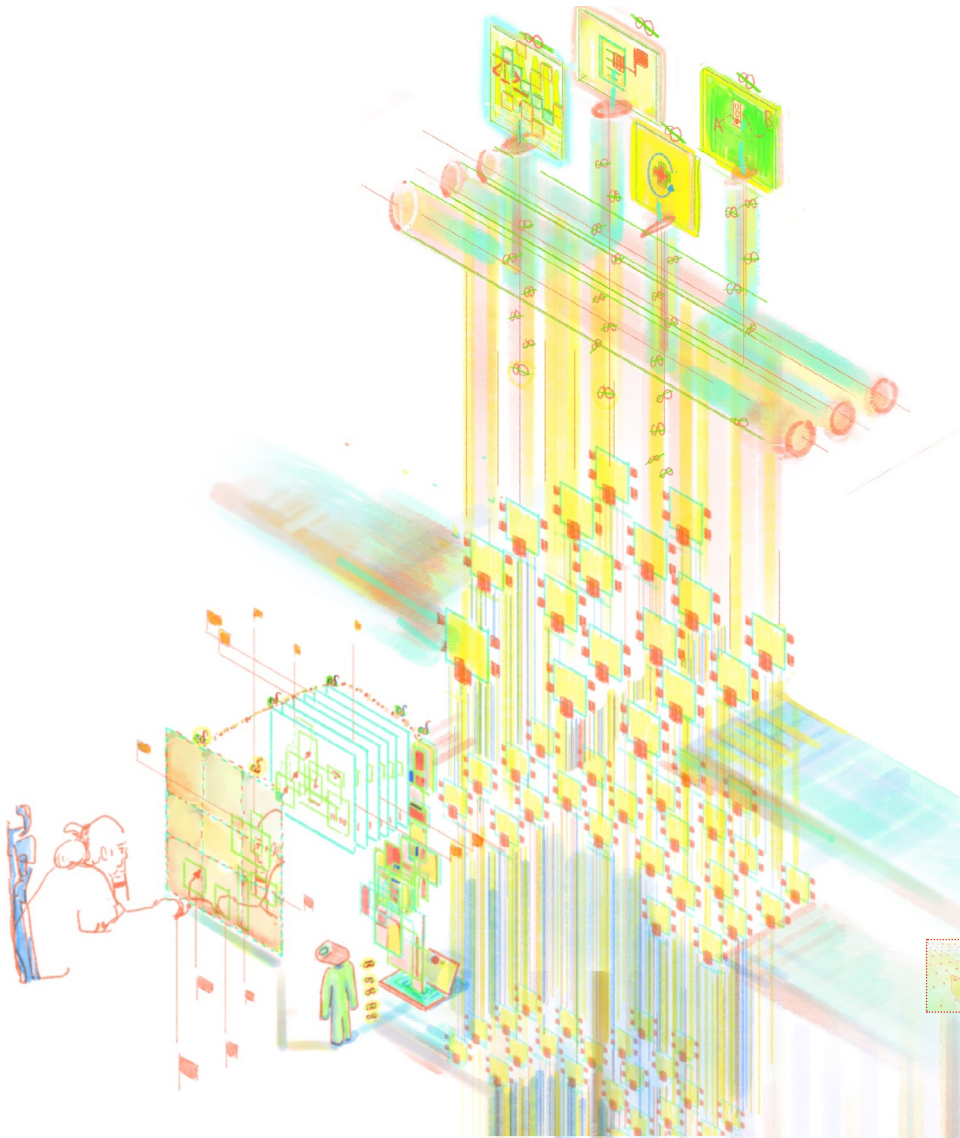


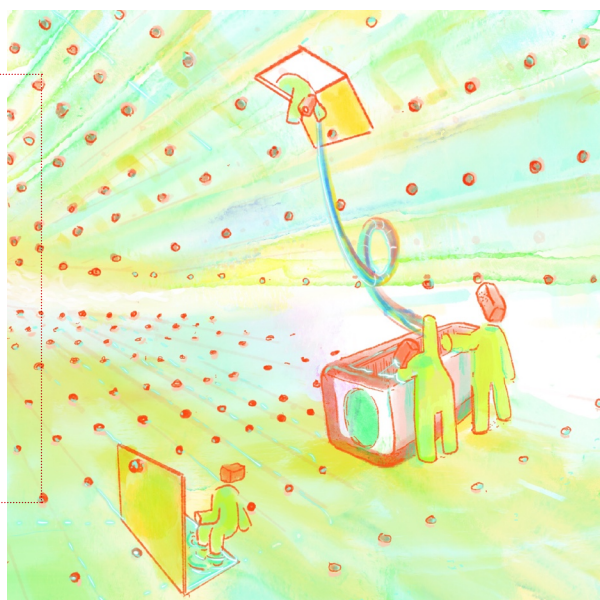
p: **Repair, Literacy and Dignity** : India has already been tagged as having the 'next Billion users' in terms of digitisation and digitalisation. Yes the contexts and linguistic terrain adds further notches to the epistemic shifts promised through machine learning and how it may help with decision making across multiple levels. However it is harder still to ignore how this diversity is also evaluated as cheap labour and precious data at the same time.

produces meaningful benefits or risks commodification and surveillance. Both organisations acknowledged that major technology companies remain the primary beneficiaries of these datasets, while local organisations face weak technical infrastructure and persistent structural barriers such as fragmented, opaque value chains and multiple intermediaries separating data workers from end clients, limiting both fair compensation and transparency.

The workshop closed with a reflection on what collective action across organisations might look like. Speakers acknowledged that no single initiative can shift the dominant logic on its own, and that a movement, involving researchers, worker organisations, civil society, and alternative producers, is needed to change how data work is valued, and to establish norms that treat data workers as co-creators of the AI systems they make possible.







Concluding Reflections and Next Steps

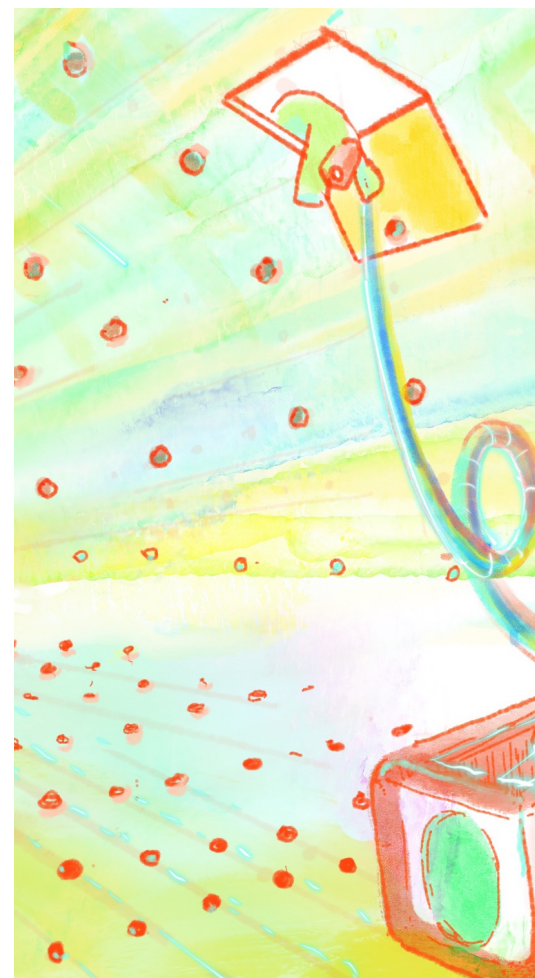
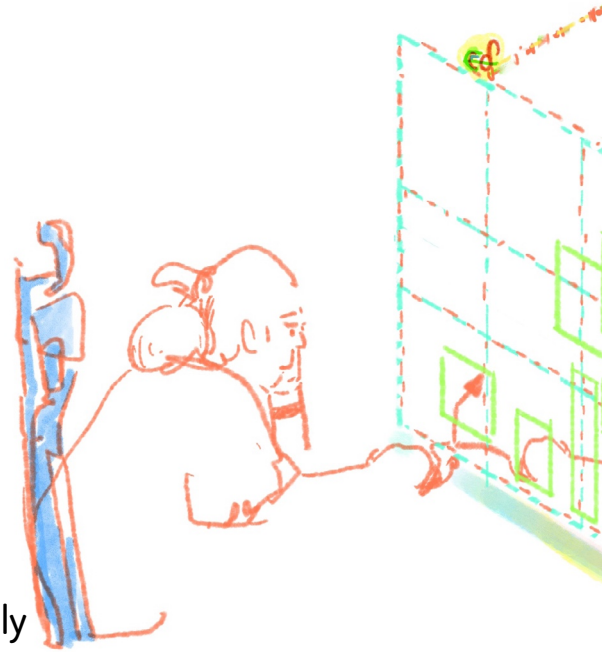
Ch 10

49

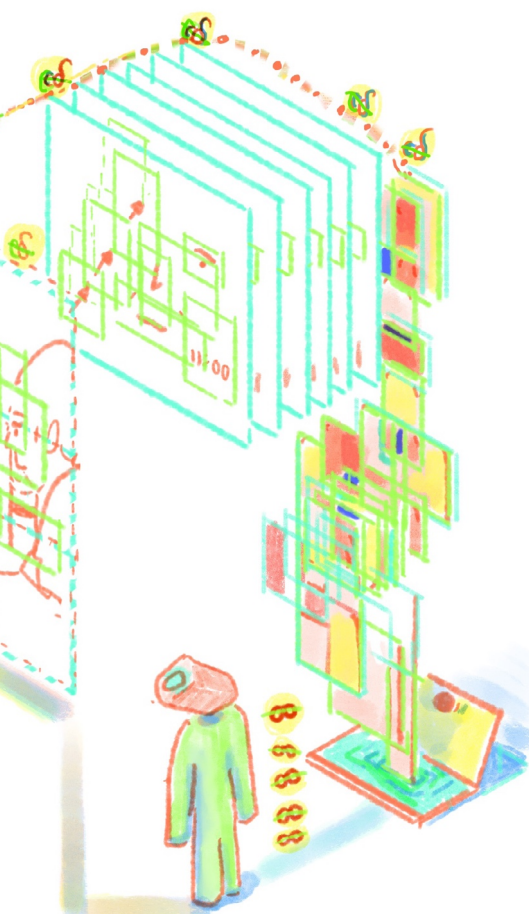
The workshop was designed from the outset as an open-ended inquiry across disciplines, sectors and geographies. Sessions moved between different modes of engagement: analytical discussion and reflective simulation, empirical research and personal testimony, critique and imagination.

A question that proved generative throughout: what is truly within the scope of 'data work'? Grouping together annotation, content moderation, CAPTCHA tasks, bureaucratic data entry, and citizen-facing digital systems reveals the shared extractive logics underlying these infrastructures. But this grouping also risks obscuring the highly uneven supply chains through which value flows across geographies.

The answer may have a material dimension and an epistemic one.



p: Beauty is in the eye of the tiger : Machine learning and its relationship to what human knowledge and sense-making might evolve into, is transitioning, in terms of lived experiences, research, governance, geo-politics and climate-crises. While dineshan's subjects work dutifully to improve what machines can learn and do, the networked nature of human behaviour and data linked to it, gives new traction to the phrase unique identities, those that are vetted or indexed by machine vision with very specific human organising ideologies, that continues to persist regardless of only human rulers or those that look to machines to play divine ruler, court jester, trader, engineer and executioner.



The material dimension concerns labour: multiple layers of subcontracting that obscure accountability, the wage suppression, the psychological toll, the absence of workplace protections across transnational supply chains of extraction.

The epistemic dimension concerns knowledge: who gets to define what counts as valid input, whose ontologies get embedded in these systems, and what gets erased in the process. Both dimensions are obscured by the dominant framing of AI as efficient, objective, innovative, transformative — a framing that depends, paradoxically, on recruiting vast numbers of people into the project of building AI while rendering their contributions, their judgments, and their contexts invisible.



Continuing and deepening this conversation requires building alliances across sites of diverse expertise and unequal stakes, remaining attentive to differentials in access, resources and institutional standing. In organising this workshop, we were confronted at various moments by our own positionality as researchers based largely in the Global North, working in collaboration with partners in India and across the Global South. Navigating these differences is likely not a “problem” to be solved at once, but an ongoing condition that must be actively worked with — through equitable partnerships, shared

agenda-setting, and sustained reflection on who benefits from the process and its outcomes.

Several directions suggest themselves. The larger histories of business process outsourcing offer lessons about the limits and possibilities of accountability in global service chains. Other sectors, for instance garment and textile supply chains, have confronted similar challenges with greater visibility and have been the site of sustained advocacy.

What can be learned from these experiences about the conditions necessary for change?

The workshop focused primarily on the labour embedded in data annotation and AI training. But the supply chains sustaining AI infrastructure extend further: to the extraction of raw materials powering data centres and devices, also concentrated in the Global South, also under conditions of severe labour exploitation and environmental harm. Finding ways to connect these struggles, without flattening their differences, remains an important avenue.

Advocacy likely looks different depending on where you stand in these supply chains, and what the target is. Technical communities, business actors, and policymakers occupy very different positions across these geographic contexts, and reaching them may require varied strategies and languages.

The next step in this conversation is to bring it to Europe — the Global North that profits most directly from these supply chains and where there may be opportunity to pursue greater accountability for the institutions that command and benefit from these infrastructures. The EU AI Act, emerging supply chain due diligence frameworks, and a growing interest in alternatives to the dominant technology paradigm, create openings to connect the concerns raised in Bengaluru to the policy processes in Brussels.



p: You made your bed, now we're all just living in it : Malaphors are strange mixes of metaphors and malpropisms. Illustrations often anchor on irony, incongruence and some familiar world rules. This report has some shades of irony when it comes to humans having to learn how to 'machine'. The fear in science fiction has always been machines taking over jobs and human life. These illustrations look at our current reality and histories that lead up to it as more stranger reconfigurations of what labour might mean in the future, and if being a citizen is tied to it in newer proportions and obscured consequences.



Appendix

The following is the list of organisers and participants who joined us for the workshop, listed alphabetically by first name and their affiliations at the time of the workshop.

Organising team

Aditya Singh (University of Edinburgh)

Srravya Chandhiramowuli (University of Edinburgh)

Jef Ausloos (University of Amsterdam)

Valentina Carraro (University of Amsterdam)

Kaustubha Kalidindi (Tattle Civic Tech)

Tarunima Prabhakar (Tattle Civic Tech)

Priyam Vadaliya (Apti Institute)

Ritvik Gupta (Apti Institute)

Soujanya Sridharan (Apti Institute)

Participants at the workshop

Sona Mewati (Indian Institute of Human Settlements)
James Oyange (African Content Moderators Union)
Aditya Ray (University of the West of England)
Riya Ravisankar (Fold Labs)
Shreya Toshniwal (Fold Labs)
Abhineet Nayyar (Centre for Internet and Society)
Chetna VM (Centre for Internet and Society)
Meera Muthukrishnan (IIIT Bangalore)
Swati Ganeshan (IIIT Bangalore)
Joyeeta Dey (Manipal Academy of Higher Education or MAHE, Bangalore)
Thomson Chakramakkil (IIT Delhi)
Uma Rani (International Labour Organisation)
Mophat Okinyi (Techworker Community Africa)
Jeevitha SatheeshKumar (JSR Script)
Sara Marcucci (AI + Planetary Justice Alliance)
Sonam Jindal (Partnership on AI)
Shruti Kaushik (India Kanoon)
Sanjana Malhotra (Columbia University)
Mila Samdub (Independent)
Chandana Dhanraj (Karya)
Nihar Desai (Project Vaani, ArtPark)
Ramkumar (Karnataka State IT/ITES Employees Union)
Sebastien Antoine (IIIT Bangalore, Maynooth University)
Sadhana Sanjay (IT For Change)
Paul Anthony George (Srishti Manipal Institute of Art, Design, & Technology)

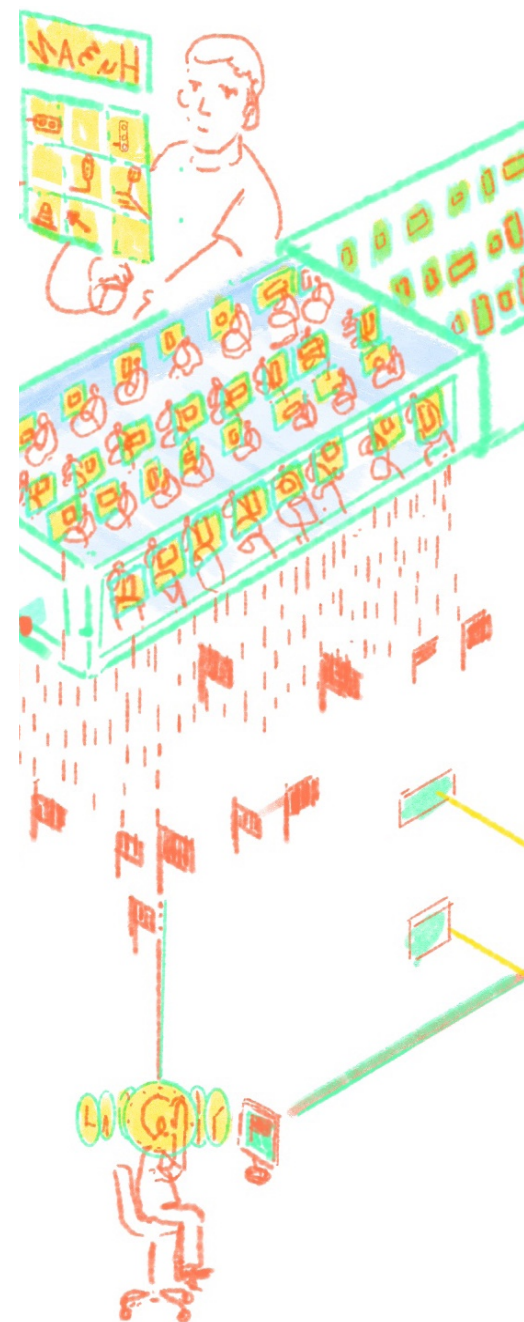
Illustrator's note:

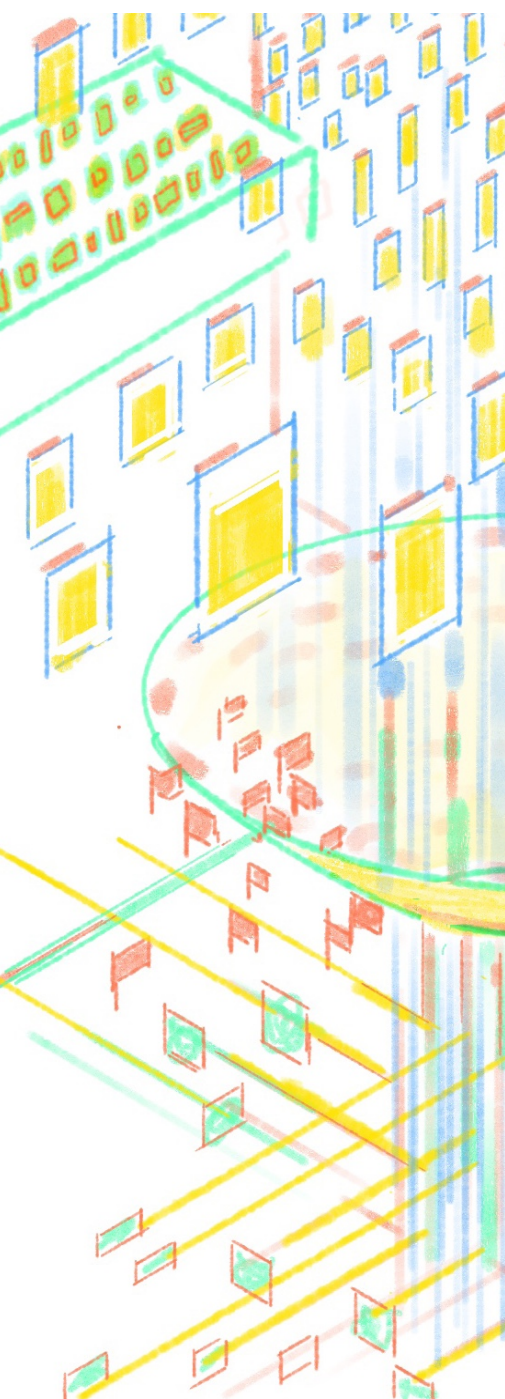
The illustrations in this narrative are an attempt to create a visual language for phenomenon around data labour, machine vision, AI and its promises, what is shown, what is seen and what is obscured or made invisible.

The actors are an extension of an ongoing mapping that started out with the Revisualising Cybersecurity project, which shows the origin stories of dineshan. A manifestation of the intermingling of corporate and state interests around platforms and citizen/ user participation. This participation being rich sources of data extraction, referred to in many ways, most recently as "Surveillance Capitalism". These relationships are further explored in the comic "dad and his invisibility mundu" (later animated as "appa and his invisible mundu") to illustrate what could happen when there is a rift between legality and justice when it comes to legislation around data and privacy.

The Ai Observatory project after, created grammar and ironed out some kinks in the world-building through how ADMS is used for governance across contexts in India.

This report extends the world of these metaphors and malaphors, by further mapping people behind the pixels that have been tagged, clicked, moderated, weighed and remashed by human hands and machine's probabilistic collages, be it generating images, text, code, or finding and mapping various world models to determine everything to faces to foetal health.





The colour red is used often to indicate something that manifests in a tangible form: subject/citizens of a state, infrastructure that allows for participation/ interaction.

The colour blue is usually an indicator of aggregated data and datasets that have been compressed or processed through manual moderation, paperwork and supervised/ unsupervised learning models.

A mix of red, blue and yellow indicates an extraction of behaviour that has been quantified in one or more ways, often under the label of “raw data”. Though while drawing these, rawness points more towards a metaphor for vegetables and uncooked-ness as opposed to referring to data as oil, from crude to pure.

Green is usually an indicator of learnt (choice-probability based) abstraction that is part of a learning model that recognises human behaviours, through translation and matching depending on the task that is being automated. Typologies for images for facial recognition, or numbers and text for welfare, license plate id'ng.

