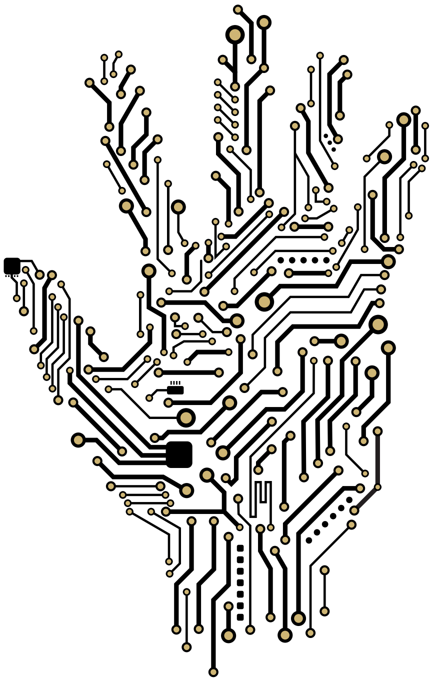
aapti institute



AI DATA LABELLING

ROUNDTABLE READBACK

This readback is based on a roundtable on AI Data Labelling, held on the 16th of June 2020, hosted by Aapti Institute and Dr. Alex Taylor, City, University of London. This is part of a larger research collaboration that intends to shed light on labelling and annotation processes, and collaboratively envision best practices for fair and equitable Artificial Intelligence. This exploratory discussion brought together startups, researchers and civil society leaders to unpack the processes and structures that are involved in data labelling or annotation work.

- BUILDING ETHICAL ARTIFICIAL INTELLIGENCE TOGETHER -

AI DATA LABELLING

The COVID-19 pandemic and lockdown are shifting the structure of markets and nature of available jobs, well-beyond the present moment. With these immediate changes and longer-term trends, it would seem timely to revisit and explore the state of technologically-mediated labour, where work is both found and delivered online. Also presenting pressing questions is the increasing promotion of AI in emerging and evolving forms of contemporary work. AI and machine learning (ML) are being touted by multinational technology firms as the backbone to future visions of work, particularly remote work. Yet many questions remain about the specific role of computational systems in these new forms of work and the shape such AI and ML systems should take. Furthermore, we are having to confront longstanding and in some cases new questions about automation in the workplace and what we imagine to be not just productive but fair and just ideas of labour.

It was with this as a backdrop that the roundtable we report on below was held. Rather than attend to the oft-hyped, idealised visions of work, however, we opted to pay attention to the work 'behind the scenes', conducted to make AI/ML work. Specifically, we wanted the primary focus of the roundtable to be on a labour that is rarely discussed and often invisible to end users, data labelling or annotation. Our interest was (and remains) in how data labelling or annotation plays a critical role in the success of the future visions of work, and what practices and structures are necessary to achieve such a success. We thus wanted to promote a discussion to better understand the conditions of this labour and both the social and technical issues that arise in enabling it.

Annotation and labelling for Artificial Intelligence (AI) offer employment opportunities for many. Across multiple sectors including e-commerce, autonomous vehicles and healthcare, AI is being promoted as an innovative solution to provide analytics, automate processes and evolve business models. However, AI is dependent on the availability of labelled and classified datasets. Without necessary meanings attached to data, machine learning models cannot be trained or function in the real-world. This critical role is played by humans who have the skills and know-how to classify new data sets, and approach problems of uncertainty and ambiguity in data and context in ways that computer models find notoriously hard to replicate.

To carry out these tasks, crowdsourcing platforms and startups often employ thousands of workers to label the datasets of text, images, video, etc. Previous work, including from one of our roundtable participants, Mary Gray, has shown the difficult labour conditions that these workers can face, surfacing not only the hidden nature of this work, but its potential for low wages, exploitation and abuse. It is clear that pathways for improving labour conditions and regulation need to be set out and where possible used to help inform national public and organisational policy programmes. In order for us to do this, we need a fuller understanding and more systematic representation of the current business models, the nature of work, and concerns of workers. Building and developing regulation for more equitable AI depends on the recognition and deeper understanding of this labor. As the future of work becomes more digital, it is imperative to understand how this work is done, who does it, and how it can be structured and performed to enhance dignity and protect rights, while enabling innovation.

METHODOLOGY

The Roundtable was divided into two sections, the first featured presentations from three startups: Playment, TaskMonk and iMerit. As limited information exists on the state of AI labelling, startups provided detailed insight into the industry and current processes. The second half provided time for researchers to discuss their work and explore areas for future research collaboration. This also enabled startups to engage in discussion around research, policy and upcoming trends.

SPEAKERS

AJINKYA MALASANE
Co-Founder,
Playment

SAMPATH HERGA
Co-founder & CEO
TaskMonk

JAI NATARAJAN
VP, Marketing &
Strategic Business Development
iMerit

MARY GRAY
Senior Principal Researcher
Microsoft Research

NIMMI RANGASWAMY
Associate Professor
IIIT Hyderabad

DING WANG
Researcher
Microsoft Research

VIVEK SESHADRI
Researcher
Microsoft Research

KALIKA BALI
Principal Researcher
Microsoft Research

SARAYU NATARAJAN
Founder
Aapti Institute

MODERATORS

DR. ALEX TAYLOR
Co-Director,
The Centre for HCI Design
City, University of London

SUHA MOHAMED
Strategy & Partnerships
Aapti Institute

PART 1: *Connecting with the Startup Ecosystem*

Startups like Playment have emerged in response to the inadequacies such as the lack of consistency in quality from legacy platforms likea Amazon Mechanical Turk. Playment seeks to democratize the availability of labelled data for other developers and machine learning engineers. Previously dominated by big tech, many companies have started to build their own models, open-source tools and platforms for annotation. Over the past decade, the nature of labelling and annotation work has evolved in complexity, type of work and expertise required. While initially labelling and annotation often involved straightforward tasks such as image classification, the work has now evolved requiring the annotation of for example 3D models, or medical imagery and scans. Though the crowd-work model is still employed, startups are leaning towards BPOs and impact sourcing firms, or in-house staff to build a secure and consistent pool of labellers that possess the requisite knowledge and skills to carry out tasks.

During the first half of the roundtable, the three startups listed below described their efforts to respond to this evolving landscape and how they have chosen to structure/implement their labelling and annotation work. They also discussed some of the challenges faced with their approaches.

PLAYMENT

Playment's mission is to expedite the AI age by availing large quantities of high quality labeled datasets to ML teams. Playment provides software and services for data labeling in computer vision space, has served 200 customers across 12 countries with more than 10,000 crowd-workers and BPO workers in India. Mercedes, Samsung, Intel, Ford and Sony are some of their key customers.

TASKMONK

TaskMonk aims to change the way annotated data is being procured by enterprises for training AI / ML Models. Within a year of launch Taskmonk has scaled to deliver 3.5 million/tasks per month for its customers serving process involving Text, Image, Video, Audio, Sensor and Geo-Spatial data.

IMERIT

iMerit leverages technology and skill development to engage women and youth from newly digital communities and develop them into expert technology workers. This advanced workforce labels, enriches, augments and audits data to shape algorithms. iMerit has over 2500 fulltime data experts with 52% women and over 80% employees from under-resourced backgrounds, with an average age of under 25. iMerit works with leaders across sectors like Autonomous Vehicles, AgTech, Medical Imaging, e-Commerce and Financial Services.

PART 1: *Connecting with the Startup Ecosystem*

"Quantity, diversity and quality of data is what matters for machine learning engineers"

- AJINKYA MALASANE, Playment

WHO ARE THE LABELLERS?

The demographic profile of data annotators and labellers vary across platforms.

- Playment labelers are typically housewives, recent graduates and retired people/veterans
- iMerit has an inclusive hiring policy – 50% of their employees are women and 80% of their workforce comes from disadvantaged communities defined based on respective communities.
- TaskMonk's annotation partners and BPOs engage workers from Tier II & Tier III cities in India

"We don't have criteria, we just need a skill set, and an aptitude test that a person needs to pass. If they passed, no matter which department qualification, age, sex, gender, we will accept them. Of course, there is definitely a working minimum of 18 years of age."
- Ajinkya Malasane, Playment

HOW IS WORK STRUCTURED AND WHAT PROCESSES ARE INVOLVED?

Data Labelling work is typically structured through crowd-work, outsourced to registered BPOs or carried out in-house with full-time employees.

- Playment adopts a hybrid crowdsource and BPO sourcing model, where labelers are onboarded if they meet basic requirements and pass an aptitude test.
- iMerit structures work through a formal, full-time employment contract and labelers are often hired from poor and marginalized communities.
- TaskMonk has created a technology platform in which their clients can carry out in-house labelling using tools or by onboarding annotation partners.

Automation is used to allocate tasks to labellers and provide them with guidance in the annotation process

- Playment's platform allocates tasks to labellers based on their degree of experience and performance ranking
- TaskMonk allocates tasks based on the workers' degree of familiarity, which also enables them to gain more domain or subject matter expertise. Alternatively, a more complex task is split up into chunks, allocated to different workers and finally stitched together once complete.

"If you want to do annotation work or computer vision, or a text transcription or text translation, we use a lot of the cloud models that are already available as a starting point and the annotation partners become the checker for that"
- Sampath Herga, TaskMonk

PART 1: *Connecting with the Startup Ecosystem*

WHAT ARE THE CHALLENGES?

1.Quality

Quality of labelled data can be improved by standards and by building better and smarter tools for annotators. Some startups like Playment and TaskMonk use a series of processes and quality standards to address this. Consensus Based, Gold Standard & 'Maker/Editor/Checker' Logic Models are three approaches for ensuring quality of data is maintained through the labelling process.

"For a data labeling provider the biggest problem to solve is quality." - AJ Malasane, Playment

"Most of the quality requirements are a little higher than what it was earlier. Most of the low hanging fruit is done. We need quality work and data annotation which works well with the managed services model"-
Sampath Herga, TaskMonk

Consensus Based Model

This entails inserting pre-annotated images in the labelling set. If labellers annotate these images correctly, it serves as a basis for how the rest of the labelled data is assessed for quality. Playment argues that this model is flawed as it is not representative or indicative of high quality labelling processes.

Gold Standard

In this model multiple annotators carry out the same labelling task, which is more effective and arguably produces better quality output. However, given that it requires several annotators, it is both more expensive and time consuming to employ as a standard.

Maker/Editor/Checker Logic Models

This model is employed by Playment and ranks annotators based on degree of expertise and skill set. Based on their proficiency with tasks, they are either classified as 'Makers', 'Editors' or 'Checkers'. Checkers tend to possess the most experience and are required to verify the work of Makers. Checkers report to their respective project managers. Based on the work required, Editors can be brought in to supplement the work of checkers. It combines the value of the Gold Standard as tasks are reviewed by at least 4-6 workers. Workers who are labelled as checkers are able to build their familiarity and expertise in annotation.

"It's a unique problem of building software and enabling human work in tandem, which makes it neither a software industry nor a complete services industry. It's a hybrid"

- AJINKYA MALASANE, Playment

PART 1: *Connecting with the Startup Ecosystem*

2. Uniqueness & Complexity of Tasks

To address new labelling demands and requirements from the industry such as labelling data in 3D models for Augmented Reality/Virtual Reality functions, startups either adopt a tools agnostic approach combined with a focus on dynamic training (iMerit) or prioritize the creation of smarter tools to enable workers with limited training (Playment).

RESPONSE FROM RESEARCHERS & SCHOLARS

Labelling image/video data is very different from labelling language or speech data for NLP models

"We have done some small data collection for Natural Language Inference recently, and it is super difficult to even design the task, let alone the issues of evaluation."
- Kalika Bali, Microsoft Research

"No two projects and jobs and data labeling activities are similar. I call this the 'how to ground truth, the ground truth' - there's no solution available on the market"" - AJ Malasane, Playment

"**I believe in making better (easier, efficient) tools to do things at scale. Tough to train and expect from 10k people. The better the tool, the lesser the variability**"- AJ Malasane, Playment

"**The difference between the person who uses a basic image editing like what we have in PowerPoint, versus somebody who knows all the tricks of Photoshop, the shortcut, the magic wand etc - that difference is substantial**" - Jai Natarajan, iMerit

3.Attracting and retaining data labellers and annotators while considering the evolving requirements and complexity of tasks

Workers that demonstrate high learnability can be guided and their capacity can be built effectively, even if they possess limited formal education and training.

RESPONSE FROM RESEARCHERS & SCHOLARS

There is often a hierarchy of expertise in the types of tasks required, some of this complexity can be overcome by skilling & training.

"As the complexity of the task increases, the complexity of the annotations required increase, you'd need a little bit more expertise than just throwing it at an unknown crowd. So, skilling and training is a very important part."-
Kalika Bali, Microsoft Research

"We are having to build a people architecture that stays one step ahead of the algorithm. This means being able to do increasingly complex work, and being able to do increasingly nuanced work and being able to provide insight to the makers of the algorithms."- Jai Natarajan, iMerit

"A lot of people have what we might call street smarts or learnability. When they're presented with the right curriculum in context, they're able to convert that into high data labeling quality."
- Jai Natarajan, iMerit

PART 2: *Exploring futures of work in AI Data Labelling*

RESEARCHERS & SCHOLARS:

DR. KALIKA BALI

PRINCIPAL RESEARCHER, MICROSOFT RESEARCH

Kalika works in the areas of Machine Learning, Natural Language Systems and Applications, as well as Technology for Emerging Markets. Her research interests lie broadly in the area of Speech and Language Technology especially in the use of linguistic models for building technology that offers a more natural Human-Computer as well as Computer-Mediated interactions. She is currently working on Project Mélange where she tries to understand, process and generate Code-mixed language data for both text and speech. Code-mixing or use of more than one language in a single conversation or utterance is a phenomenon that is observed in all multilingual societies.

MARY L. GRAY

SENIOR RESEARCHER, MICROSOFT | E.J SAFRA CENTRE FOR ETHICS FELLOW | FACULTY AFFILIATE, BERKMAN KLEIN CENTRE FOR INTERNET & SOCIETY, HARVARD UNIVERSITY | FACULTY - SCHOOL OF INFORMATICS, COMPUTING & ENGINEERING, INDIANA UNIVERSITY

Mary, an anthropologist and media scholar by training, focuses on how everyday uses of technologies transform people's lives. At MSR, Mary works on the social impact of digital labor through the case of on-demand labor. In May 2019, with computer scientist Siddharth Suri, Mary published five years of collaborative research in the book, Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass published by HMH Books. Their research team combined methods from anthropology and computer science, to amass the largest data set about on-demand work ever collected. They conducted hundreds of in-person interviews and more than 10,000 surveys of workers in the US and India.

DR. NIMMI RANGASWAMY

ASSOCIATE PROFESSOR, IIT HYDERABAD

Nimmi is an Associate Professor at the Kohli Centre on Intelligent Systems, Indian Institute of Information Technology, Hyderabad. She brings an anthropological lens in understanding the impacts of AI research and praxis. Nimmi is also an Adjunct Professor at the Indian Institute of Technology, IIT, Hyderabad where she teaches courses at the intersections of society and technology. Previously, Nimmi was a senior research scientist and led the Human Interactions research area at the Xerox Research Center India. Nimmi previously worked with Microsoft Research, with a focus on a combination of theoretical analysis and ethnographic field research to understand technology use in developing countries.

DR. VIVEK SESHADRI

SENIOR RESEARCHER, MICROSOFT RESEARCH

At MSR Vivek is broadly interested in designing efficient computer systems, and more recently, in developing technologies for societal impact. One of his current areas of focus includes Project Karya, a new platform to provide digital work to rural communities. The word "Karya" literally means "work" in a number of Indian languages. Project Karya enables rural communities to participate in crowdsourcing. Vivek primarily works with the Technology for Emerging Markets group at Microsoft Research India. He received his Ph.D. in Computer Science from Carnegie Mellon University. His Ph.D. thesis was on new memory abstractions for efficient memory systems, with specific focus on DRAM-based main memory.

PART 2: *Exploring futures of work in AI Data Labelling*

While crowd work and content moderation processes have been significantly documented, limited research exists on what this may involve for data labelling processes, necessary for building AI. Often rendered invisible, the labour behind these data labelling processes are more frequently being carried out by workers in the global south. These experiences are further mediated by dimensions of class, gender, language proficiency and technical expertise. This section provided researchers with the opportunity to share learnings from their research, explore avenues for collaboration, and raise questions. While this section of discussions was limited by time constraints, key themes and issues raised by researchers set the stage for further discussion and areas for future research.

LEARNINGS FROM EXISTING RESEARCH:

KALIKA BALI

Prioritize and value the inclusion of women and people from underrepresented and marginalized communities

There are often assumptions around people who carry out platform based work and limited understanding of how this work can improve or perpetuate existing patriarchal structures. Further research needs to be done on gender and its relation with social capital

"Is it really true that women are able to take their own decisions because they're earning money? Is it just that the money is just handed over to the head of the family and he takes all the decisions? - Kalika Bali

MARY GRAY

Acknowledge that data labelling/annotation is not 'click-work' or a niche job

The processes, expertise, time and energy spent on these tasks should not be rendered invisible but instead recognized so we can build better support systems.

"Imagining we are somehow reducing what it is that people are doing fails to see just how incredibly cognitive cognitively challenging it is to be able to be constantly offering that very human capacity for spontaneous responsive insight" - Mary Gray

PART 2: *Exploring futures of work in AI Data Labelling*

Foster a culture of collaboration between workers both on and offline

Whether or not this is encouraged by platforms, workers typically find peers and form networks that they then use to ask for help and collaborate.

"We accomplish what we accomplish in our day to day employment, through groups of people coordinating and collaborating" - Mary Gray

Re-orient workplace environments and consider what can be done differently to strengthen well-being, conditions and the ecosystem in which workers operate

How can we re-shape the Future of Work away from white, neo-liberal conceptions of productivity and accomplishment?

"What are ways in which we would recognize what are people trying to get out of these forms of employment? How does it speak to them meeting needs that are not met in the way we try to formally organize employment now?" - Mary Gray

DR. NIMMI RANGASWAMY

Build relevant methodological approaches to enable ethnographic study on the Future of Work

From the perspective of anthropology, what is the methodology and approach to adopt when studying automation and the impact of artificial intelligence?

DR. VIVEK SESHADRI

Explore how digital work (annotation and labelling) can be made a more accessible livelihood opportunity for rural communities

Through Project Karya, Vivek works with rural communities who help label and digitize regional language and speech sets to build natural language processing models for the next billion users in India.

"Language based skills are one of the strong skills of people in rural communities. What we essentially want to do is take the wave of digital work that is imminent in language based machine learning models and build a platform that can take that work to the communities which are skilled in those types of tasks." - Dr.Vivek Seshadri

PART 3: *Learnings for Industry & the Way Forward for Research*

KEY LEARNINGS FOR INDUSTRY:

Building a community of workers that is highly skilled, motivated and incentivized to continue learning is a value in itself as it enhances worker well-being; it is also associated with better quality output.

"...we wanted to build a strong community, which doesn't fight against each other for the monetary benefits but likes working together to make the end job successful" -
AJ Malasane, Playment

Diverse workforces are good for business and can improve retention and commitment to learning and building skill sets for success

"80% of who we hire are actually from very poor backgrounds, who are completely locked out. What we get is phenomenal retention, phenomenal motivation. We get people that for whom this work is transformational in their communities" -
Jai Natarajan, iMerit

RESPONSE FROM RESEARCHERS & SCHOLARS

It is critical to consider how to shape supportive peer networks for data labellers as this is instrumental in determining work-experience and subsequently in driving better quality output

"There's so much energy spent on how to get high quality out of people when the answer has been there all along. It's you care about them, you care about them having an environment in which they can do good work, you get high quality data."
\- Mary Gray

PART 3: *Learnings for Industry & the Way Forward for Research*

APPROACHES TO IMPLEMENTATION:

TRAINING: Investing in dynamic training and skilling of data labelers enables them to rise up a skill curve

'Super labelers', 'Checkers' and formally employed data labelers often deliver higher quality output as a result of familiarity with processes and learned expertise.

"there was a time we had 500,000 annotators on the platform. We kind of squeeze that down to really high quality individuals who take this as for their living, their incentives are in line. They can spend time they understand higher levels of concepts" - Ajinkya Malasane, Playment

TRAINING: Intuitive training designed and lead by domain experts can supplement labellers with limited expertise and enable more inclusive hiring policies

iMerit relies on in-house domain experts who design curricula based on labellers pre-existing knowledge. As a result, labellers with limited educational qualifications are able to effectively annotate medical imagery and carry out other complex tasks.

"A lot of people have what we might call street smarts or learnability. And when they're presented with the right curriculum in context, they're able to convert that into high data labeling quality"- Jai Natarajan, iMerit

"We have two PhDs in linguistics. They sit down with the customer and they unpack the customer's jargon rich domain and they convert that to a data labeling model" - Jai Natarajan, iMerit

HUMAN RESOURCES: Building sustainable career trajectories for workers by investing in their development is critical for enhancing livelihood opportunities and can also create positive business outcomes

Investing in workers and their career development in a way that facilitates growth and mobility within the company can have an impact on the individual and their respective communities. As a company that is dedicated to creating lasting impact, iMerit claims that this approach does not hinder performance, cost, quality or agility compared to its industry competitors. In contrast, they argue that the next upcoming trend in the industry is to see the human or 'worker' in the loop even after the labelling/annotation is deployed.

"We've done surveys of our employees in the neighborhoods, the communities that we operate in...We find that over 80% of them [workers] pay back in the community, they educate a sibling or a family member or in turn they pursue their own education in night college. They build things like sanitation in their communities and in their homes. We find that virtuous cycle really working."- Jai Natarajan, iMerit

"The future of work is not a what, but a how"

- JAI NATARAJAN, IMERIT

AREAS FOR FUTURE RESEARCH AND POLICY CONSIDERATIONS:

The areas for future research and policy considerations we present below have been drawn from the roundtable presentations and discussions. We've aimed to identify areas that we feel demand future study but that might also have impactful implications for the varied stakeholders and actors. We have broken down the areas into three groupings of issues. These are intended as an aid for organising the content, and we recognise the issues likely intersect in multiple ways.

TECHNICAL AND SERVICE ISSUES:

How can we best support an evolving industry in which labelling is moving from simple tasks to more complicated skilled/knowledge work?

The changing nature of the tasks and activities coming from data labelling customers is presenting new challenges in terms of measuring both output and quality. This presents more than merely technical hurdles. It raises issues around how work should be judged fairly; the responsibility the industry should take in supporting knowledge work; and how such work should receive recognition. It also speaks to a wider set of concerns surrounding the ways in which skilled labour is being increasingly parameterised and valued in purely monetary terms. As the data labelling industry moves towards measuring output and especially quality, further research is needed to explore how best to recognise the labour force, and how to do more than provide individual financial reward, and consider the prosperity and sustainability of communities.

How should we make sense of a focus on platform building in provisioning data labelling services?

All the startups contributing to the roundtable discussions spoke of their efforts to build bespoke platforms. These focused on providing access to labellers, quantifying labour and supporting/measuring quality. A variety of questions arise around this, including how an ecosystem (or competing landscape) of heterogeneous platforms should be understood, mapped and regulated; how such technical platforms distribute and organise labour and skill; and where these systems locate control and authority?

SOCIOTECHNICAL ISSUES:

How can we begin to understand the role of language in data labelling and specifically the role of labelling speech and language use?

With the increasing trend towards more complex labelling activities to support AI systems, some platforms are trying to cope with complex language labelling tasks. This has shown the underlying challenges faced in this ostensibly technical work. Language is being shown to be unlike, for example, image labelling, as it requires a much more sophisticated understanding of the structures of language use and contexts in which language is used. Furthermore, we need to recognise there will be a variability of language use by labellers, one that intersects with social/cultural dimensions such as education, class/caste, geography, coloniality, etc. (see below). Language understanding is seen as an important area in AI and machine learning, but through the work of data labelling it reveals the need for significant future study.

How can labelling platforms enhance and build on collective/communal labour?

Labelling and annotation platforms currently apply an individual user model, aiming to maximise the labelled outputs individuals produce. As the industry moves towards quality as a metric, a greater understanding of collective or communal work is needed. From cases such as micro loan schemes, we know that many of the people doing labelling work live in and draw on communal skills and knowledge. This invites future work understanding the relations between labelling work and collective skills, and where platforms might emphasise not just drawing on but sustaining communities.

STRUCTURAL ISSUES

What are the ways in which gender intersects with data labelling work? What are the opportunities, outcomes and impact on women's lives and social capital?

One of the startups described its efforts to hire women, with a workforce made up of 60% women. This was presented as an emancipatory venture, offering women a source of income, and greater independence and skill/expertise. Beyond these benefits, we also need to consider how such changes might reconfigure wider social dynamics, ranging from changes in the domestic realm to the wider labour market.

How should we account for other wider structural issues that intersect with labour, such as race and coloniality?

As in other parts of the technology sector, what is seen as the laborious work of innovation is being exported to regions in the Global South, such as India, Vietnam, Kenya and Uganda. As well as issues of recognition and reward, this raises questions about the role of race and coloniality, and, specifically for labelling, how histories of cultural integration (and power and subordination) play into the work of category making.

What is required at a systemic and institutional level to enable justice for workers? How can we frame regulation or policy keeping in mind the diversity of lived realities, experiences and differences in how work may be structured by platforms?

We must ask how equity and justice are achieved for the workforce with the labelling industry becoming a substantial sector in global economies, particularly in emerging markets. The structures at play in sustaining the industry are built on logics that rely on exploitation, and alongside it the datafication of labour and digital surveillance. Moreover, workers who are an essential part of making AI have no role in governance. Amidst these conditions, we need to explore how the conditions for equity and justice are afforded and ensured.