What is Worth Seeing? Evaluations and Negotiation over Classification Rules in an AI-Based Medical Imaging Device

Joaquín Yrivarren

Barcelona Science and Technology Studies Group (STS-b), Department of Social Psychology, Universitat Autònoma de Barcelona, Spain/ joaquin.yrivarren@autonoma.cat

Miquel Domènech

Barcelona Science and Technology Studies Group (STS-b), Department of Social Psychology, Universitat Autònoma de Barcelona, Spain

Abstract

We examine the implementation of an AI-based imaging system in a clinical microbiology laboratory, focusing on the negotiation between AI developers and laboratory members over classification rules to be embedded into the system prior to its deployment in a clinical setting. The negotiation was triggered by a discrepancy identified during a performance evaluation, when the AI detected microcolonies that laboratory technicians had overlooked. While STS research has explored how medical imaging technologies, once in use, shape healthcare professionals' diagnostic knowledge, we adopt a technology-in-the-making perspective to analyse how the negotiation process shapes microbiologists' visual practices even before the system becomes part of routine laboratory work. Drawing on ethnographic fieldwork in the supplier company's R&D department, we analyse how different evaluations of the visibility status of microcolonies emerge through product customisation and developer–client interactions, and how contrasting viewpoints of what is worth seeing are negotiated into agreement. We argue negotiation over technological capabilities shapes visual routines by unfolding as a process of coproducing a socio-natural reality. As a contribution to ongoing STS debates, we introduce the term 'image reversibility' to capture a mode of problematising conventions underlying AI system development. We also discuss the implications of mobilising the notion of 'artisanal intelligence' to encapsulate an implementation style that favours adaptability by embedding specific laboratory practices within the system.

Keywords*:* Knowledge Co-production, Visual Practices, Full Laboratory Automation, Artificial Intelligence, Health Technologies, Microbial Identification.

Introduction

Over the past few decades, the potential of digitalisation, robotics, and AI to play a transformative role in healthcare systems—particularly by enhancing operational capacity and efficiency—has attracted sustained attention in business and political circles. At the same time, academic debates have increasingly focused on the challenges associated with using robots and AI in practices of care. In this context, Science and Technology Studies (STS) have offered an alternative perspective to the essentialist narratives associated with the so-called Fourth Industrial Revolution (Schiølin, 2020), narratives that have strongly influenced managerial discourses on technological change in healthcare (WEF, 2019).

From this vantage point, some STS scholarship has highlighted a stark contrast between an ethics of care—situated and relational—and a technocratic mindset—objectifying and entrepreneurial (VallèsPeris and Domènech, 2020). This contrast can be found underpinning more critical, and at times pessimistic, interpretations of AI in biomedicine. For example, Scherz (2019) warns of a declining confidence in human judgement as AI's 'correlation is enough' paradigm gains prominence. This tension has been described as a confrontation between two epistemic cultures: one 'dry,' centred on computational models and calculation, and the other 'wet,' characterised by contingency and close engagement with biological specimens.

Despite the richness of these discussions, one aspect has received comparatively limited attention: the practical work of implementing robotics and AI applications in clinical settings. Here the two cultures, while contrasting, converge through commercial transactions in which such applications are refined and adapted to customer needs. This paper focuses on this convergence. Specifically, we examine a negotiation between the developers of a proprietary imaging system (IS) and members of a microbiology laboratory that had acquired the system and sought to integrate it into its daily practices. The IS integrates digital image analysis of culture agar plates with expert rules that connect semiquantitative (e.g. bacterial load) and qualitative (e.g. morphology) assessments of colony growth to clinical and demographic data. On this basis, the system provides users with plate classification results grounded in presumptive microbial identification, which must subsequently be reviewed and confirmed by laboratory technicians.

The negotiation analysed here emerged from a discrepancy encountered during the implementation of the IS. The imaging algorithms detected microcolonies on plates that had previously been processed in the laboratory, classifying them as infinitesimal growths requiring further review. By contrast, laboratory technicians had overlooked these infinitesimal entities during routine visual screening. In this instance, the IS registered forms of growth that exceeded human visual detection. According to the AI developers—who had extensive experience working with laboratories worldwide—such growths would typically be disregarded within an automated laboratory workflow.

In practice, both the AI developers and the laboratory workers, drawing on their respective experiences, were of the same mind about the fate of these microcolonies, even though the imaging algorithms and classification rules did not. Despite that tacit consensus, considerable time was devoted to discussing how best to handle these infinitesimal entities and to determining which classification rules should be encoded in the system to manage them. What was it, sociologically speaking, that they were seeking to settle through this negotiation?

In addressing this issue, we pursue two related aims. On the one hand, we seek to expand current understandings of the adoption of AI medical devices in clinical settings (FDA, 2021), with particular attention to AI-based *imaging* systems designed for pathogen identification. Despite their growing presence in laboratories, such systems have received relatively little attention in specialised journals compared with other AI applications in bacteriology (PeifferSmadja et al., 2020). Yet, for both expert microbiologists and social scientists alike, the introduction of AI-based imaging into microbiology laboratories raises several foundational questions: what counts as knowledge, and under what conditions do images become reliable diagnostic evidence (Aeffner et al., 2019; Lalumera and Fanti, 2020)? Drawing on the experiences of both developers and clients, we shed light on the practical challenges of implementing AI-based imaging as a knowledge infrastructure, thereby contributing to existing studies on stakeholder concerns and expectations (Drogt et al., 2022; Paranjape et al., 2021).

On the other hand, we aim to contribute to STS scholarship on imaging technologies. As we discuss in greater detail below, much of the existing literature has examined how such technologies, *once in use in clinical settings*, mediate healthcare professionals' ways of seeing and knowing. Given the specific features of our case—most notably that the parties were in the process of finalising a commercial transaction—we instead step back to ask: how does the negotiation process shape microbiologists' visual practices before the system is fully integrated into the laboratory workflow?

Brief overview of our conceptual repertoire

Building on ethnographic fieldwork that prioritised thick, infrareflexive description over detached theorisation (Latour, 1998; Ponterotto, 2006), this study develops a modest conceptual repertoire—aligned with STS commitments to conceptual minimalism (Becker, 2011; Gad and Ribes, 2014)—in close dialogue with empirical materials. The text is organised into three sections1, within which a set of concepts is unfolded to help make sense of a range of observations and, more importantly, to situate these observation–concept articulations within ongoing STS debates on imaging systems in-the-making.

In the first section, we examine how the IS's software modules are customised, situating the alignment of AI technologies with laboratory practices within broader discussions on groundtruthing, databasing, and algorithmic training (Jaton, 2017, 2021). Drawing on Jaton's notion of 'hesitation', we conceive it as a key mediating force shaping evaluative processes related to artificial vision calibration, algorithmic accuracy, and overall system performance. In our case, hesitation emerges through and enables comparisons between AI-driven and human visual skills in company–customer interactions. In the second section, we analyse how negotiations are structured by mobilising the notion of 'worthiness', understood as the capacity of an entity to warrant collective attention (Argudo-Portal and Domènech, 2022). Drawing on Callon & Muniesa (2005), we show how questions of worthiness open a space for weighing alternative states of the world—specifically, whether microcolonies should be included in or excluded from clinical sight. We further employ the concept of 'domain' (Ribes et al., 2019) to interrogate assumptions about what counts as the real world in AI implementation, as well as the challenges involved in formalising laboratory practices. The third section addresses disputes over the visibility status of microcolonies. To capture laboratory members' contestation of visual conventions embedded in the IS, we introduce the concept of 'image reversibility'. Building on Latour's notion of 'modulation' (Latour, 1987, 1999) and Collins' 'pain of regress' (1987), this concept highlights how images and classifications are reopened and traced back to their conditions of production during negotiation. It also foregrounds the difficulties involved both in translating everyday laboratory practices into logical rules and in articulating, in simplified terms, the criteria guiding algorithm design.

Finally, we mobilise the notion of 'artisanal intelligence', which we came across by chance during feedback sessions on a draft of our manuscript. At first glance, the term underscores customisation and adaptability as a valuable sociotechnical effort through which heterogeneous laboratory protocols are embedded into the system's configuration. More fundamentally, however, by ironically reworking the term 'artificial', artisanal intelligence operates as a pointed allegory for a constitutive tension of AI-based imaging devices: although their calculative capacities enable the identification of infinitesimal entities that even an expert eye might overlook, this very abundance of visual detail can undermine the system's practical usefulness in situated clinical settings. Viewed in this way, artisanal intelligence can be understood as a productive sociological lens for apprehending the practices and interactions through which certain realities are co-produced as worthy of seeing and knowing.

Our ethnographic fieldwork

The case foregrounded here emerges from seven-month ethnographic fieldwork conducted at a

European company specialising in workflow automation within clinical microbiology laboratories. Over the past seventeen years, the company's R&D teams have developed robotics and software modules integrated into a broader automation ecosystem, with the IS constituting a prominent component of this infrastructure. Much of the fieldwork focused on observing the Imaging Analysis team, which at the time comprised ten young developers working in computer vision, AI, and data analytics.

Participant observation was undertaken by one of the authors (JY), who had fulltime access to offices and common areas three days a week. Access was facilitated by the company's CEO, and the role of the field researcher was clearly communicated to all participants from the outset. Ethnographic material was collected through daily observations recorded in a field logbook, the compilation of official and grey literature, and openended interviews. Throughout the study, we ensured the anonymity of all participants and safeguarded sensitive and proprietary information.

Our ethnographic approach draws on the theory–method package commonly associated with actor-network theory (Silvast and Virtanen, 2023; Sismondo, 2010). Accordingly, we committed ourselves to tracing R&D teams' activities in their everyday working situations without imposing predefined sociological categories, thereby increasing the likelihood of accounting for a wide range of elements—both human and nonhuman—and their associations (Callon, 1986, 1993; Latour, 1987, 2005; Law, 2009). This flexible approach has proven valuable for analysing the sociotechnical complexities of robotics and AI development and imaginaries (VallèsPeris, and Domènech, 2021). In our case, following associations enabled us to reconstruct a relatively unusual aspect of AI developers' work: direct negotiations with clients aimed at refining the product's application in laboratory routines2.

Imaging devices: From stable actors to objects of negotiation

Since healthcare became increasingly technologised in the mid-twentieth century, a powerful narrative surrounding imaging devices has gained momentum. Beliefs that such devices inherently bring progress—thanks to their purportedly transparent access to nature—have reinforced their status as infrastructures of authoritative knowledge (Harris, 2011; Joyce, 2005). A similar narrative appears to underpin contemporary optimism about AI-based imaging systems. Algorithmic calculations are often regarded as accurate representations of nature that promise to enhance healthcare professionals' vision; however, this view largely obscures the considerable sociotechnical work required to render such calculations possible.

STS scholars have challenged this narrative by examining how imaging technologies—particularly X-rays and MRI—have become deeply embedded in diagnostic practice. For example, Prasad (2005) has highlighted the emergence of a new visual regime, termed 'cyborg visuality', resulting from the production and interpretation of computer-based images generated by MRI, together with other cross-referential inscriptions used in radiological differential analysis. Within this visual regime, "not only have the roles of humans and machines in the production of images changed but so have the nature and status of the image itself" (Prasad, 2005: 309).

In turn Joyce (2005) has argued that anatomical images must be inscribed within a sociotechnical assemblage to produce medical knowledge. From this perspective, medical knowledge is understood as mediated by concrete—though not always visible—interpretative practices, such as testing, reporting, and reading, which transform images of the human body, processed by imaging devices, into clinical evidence. Similarly, Van Baalen et al. (2017), in their analysis of the hermeneutic strategies through which a shared diagnostic vision is established, argue that imaging devices actively shape the epistemic domain of health professionals "by shaping what counts as evidence for specific diagnoses and by shaping classificatory structures and treatment regimens for diagnoses" (Van Baalen et al., 2017: 955). Analogously, Drogt & Milota (2021), by tracing cycles of visibility and invisibility surrounding a suspected pathology, examined how digital technologies shape physicians' ways of seeing.

A common thread running through these studies is their emphasis on the mediating role of imaging technologies in diagnostic practice, as well as on the interpretation of images as a situated social activity (Joyce, 2006). This perspective resonates with key tenets of constructivist approaches to scientific knowledge, which foreground representational techniques and laboratory practices in explaining how knowledge is produced (Fujimura, 1992; Latour, 1986; Latour and Woolgar, 1986; Lynch, 1988). However, by focusing primarily on how practitioners' ways of seeing both shape and are shaped by medical images, much of this literature implicitly portrays such images as stable outputs of imaging devices, whose value in everyday diagnostic routines is often taken for granted. In other words, these devices tend to be treated as stable actors within the diagnostic process.

We cannot adopt this assumption. In our case, some of the IS's technical components were still being refined, and clients' attachment to the system was still evolving. Based on our fieldwork, we therefore adopt a different point of departure—imaging technology in the making. Thus, we examine the negotiations through which actors refine and allocate imaging software capabilities to meet laboratory requirements and expectations. From here, we ask to what extent negotiation shapes microbiologists' visual practices prior to the IS's integration into routine laboratory work.

In this respect, we align with Yoxen's (1987) argument that the link between image generation and consensus over the value of images is central to understanding the consolidation of imaging technologies in medical settings. A similar point is advanced by Coopmans (2010) in a case study of demonstrations of imaging software developed by a small British university startup seeking commercial applications. By analysing how the software's capabilities were presented to prospective clients, Coopmans shows that the reliability of representational content is co-produced alongside its commercial value (e.g. its profit-making potential). Although our case differs in that it does not centre on a demonstration, we share Coopmans' interest in imaging software in-the-making, particularly in situations where representational capacities and customer attachment are simultaneously at stake.

Shaping a medical imaging device

We now turn to key events related to product customisation, with particular attention to a series of evaluations concerning three main components of the IS: artificial vision, image analysis software, and the expert system. In doing so, we emphasise that the AI's capacity to detect bacterial growth and classify digital images is not fixed; rather, it is

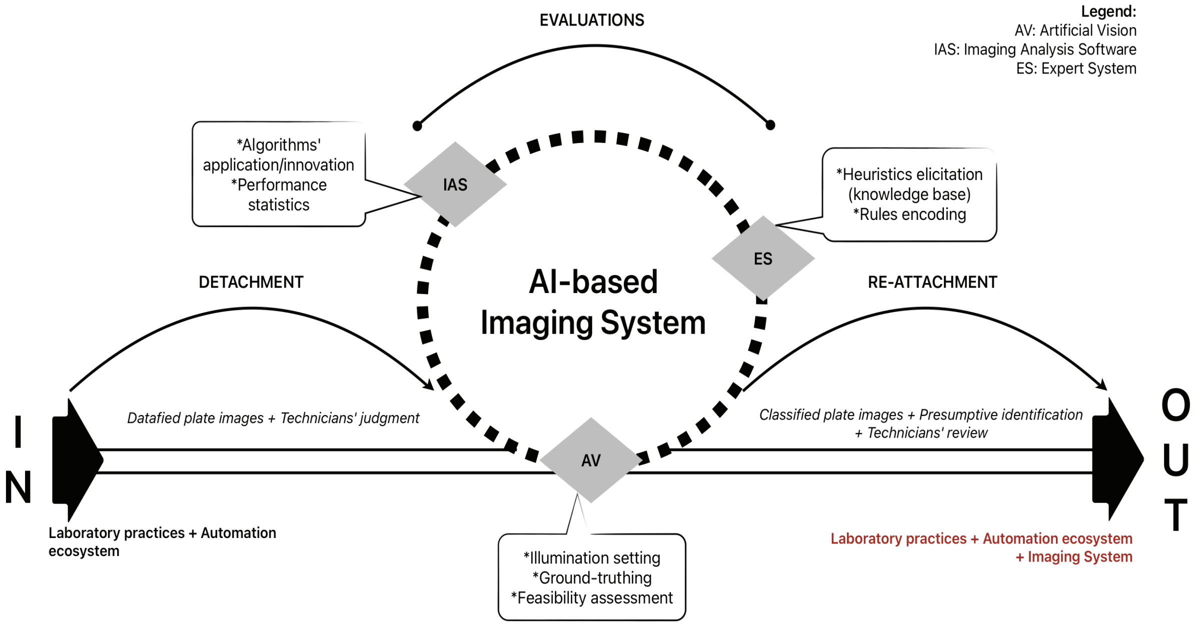


Figure 1. Implementation process of the Imaging System.

subject to deliberate fine-tuning driven by interactions with clients and by comparisons with laboratory technicians' judgments. Figure 1 illustrates the processes of detachment, evaluation, and reattachment that digital images undergo during the implementation/customisation of the IS.

Artificial vision

The first technology is artificial vision, which is used to extract information from digital images. Artificial vision can be understood as an instance of what Lynch (1988) has termed an 'externalised retina', referring to the ensemble of operations and techniques through which specimens are rendered visible, calculable, and thus transformed into docile objects of knowledge. Achieving this requires a dedicated infrastructure. In our case, it consists of an optical module integrated into the automation ecosystem—already installed in the client's laboratory—which captures images of culture plates at different incubation times and under varying lighting conditions.

This module constitutes the foundation of the IS, as it ensures a continuous supply of stable, high-quality images that are subsequently processed by the image analysis software (see Figure 1, from IN to AV). The set of operations involved in producing these data-driven images is collectively referred to as *pre-processing*—a core component of artificial vision. As a result of pre-processing, the experience of viewing plate images differs markedly from that of observing physical plates. Crucially, these images are produced not only for laboratory technicians but also for algorithms.

> This is all about getting images ready so algorithms can process them. [E41, 1:11:24]3.
> The camera snaps a picture, but it's not the one you see on the screen. First, some calculations need to happen. For instance, when the camera takes a photo of the plate, you won't see everything around it in the interface. The images you see are different from the original shots. [E13, 51:27]

A key operation in calibrating artificial vision, referred to as image learning, involves evaluating image quality to determine optimal illumination settings. This evaluation takes place during the installation of the automation ecosystem and combines choices made by AI developers to ensure the proper functioning of the system with input from senior laboratory members, who define routine protocols—particularly regarding plate types, sample materials, microorganisms, and incubation times. Because final validation of image quality lies with senior laboratory members, customers *ultimately* participate in shaping the computational capacity of the imaging algorithms that process these images.

Over time, AI developers have standardised a set of illumination settings that function as a company asset, condensing expertise accumulated across diverse laboratory protocols. Yet such standardisation does not resolve the *contingencies inherent in practices of externalising vision*. In the case examined here, the laboratory used a plate type composed of two distinct agar surfaces, one of which had never been configured in the IS before. This unfamiliarity prompted concerns about the system's accuracy, voiced by the imaging team manager:

> We're worried about the challenges in spotting white colonies on a white [opaque] background […]. Even when observed with the naked eye, it's tough to tell those growths apart. We cannot guarantee how well it will perform, so it'll take a few weeks to assess the feasibility. Our suggestion is to switch out the plate for a clearer one. [24.3, I, p. 106]4

The presence of plates with such characteristics in laboratory protocols led the team to conduct a *feasibility assessment* to determine whether, under a specific illumination setting already used in the laboratory automation ecosystem, these plates would generate sufficient agar–colony contrast for the imaging algorithms to operate effectively. We treat this feasibility assessment as an instance of what Jaton (2021: 6) refers to as *ground-truthing practices*: "the diverse set of actions aimed at verifying the accuracy of a computerized method of calculation". As the term suggests, the outcome of such verification is the delineation of what comes to be designated as ground truth. During a formative training course delivered by the company, an experienced AI developer described ground truth in explicitly constructivist terms: "A level of reality defined either by an expert or by machine analysis, which is then accepted as truth during the training and validation phases of a model." The underlying idea, then, is that evaluating the technical reliability of the IS's externalised retina amounts to constructing a reality-as-ground-truth upon which algorithms can subsequently operate.

This line of argument echoes William James's (1907) pragmatist understanding of truth (*veritas*), not as a property inherent to an idea but as something that happens to it—truthfulness emerges through processes of verification. Building on this perspective, Jaton draws on James's notion of 'morality' to introduce a further analytical nuance for understanding how accuracy is verified. Here morality is conceived as *the lived experience of hesitation surrounding genuine options and their potential consequences.* Consequently, the morality of groundtruthing resides in the extent to which hesitation permeates practices such as problem definition, datas et construction, and labelling. Put more practically, this hesitation materialises in a series of questions: Can the IS's performance meet customer demands? Must illumination settings be recalibrated? Can already developed, black-boxed algorithms detect colony growth on opaque agar, or should they be enhanced?

The feasibility evaluation can also be understood as a multilayered task involving both resource allocation and the fostering of customer collaboration. Accordingly, AI developers began collecting plate images (500–5,000), along with anonymised demographic and clinical data, to assess whether already developed algorithms could be applied to the plate type used in the client's laboratory. In contrast to many data collection practices that have characterised the AI field over recent decades, data collection here was explicitly consensual. Both parties agreed to cooperate to ensure continuity in the lifecycle of digital images. As one developer noted: "The lifecycle of the images, after the lab technicians provide their results, continues when we download them." [24.3, II, p. 111].

In this process, images undergo a crucial form of *detachment*, losing their connection to real, flesh-and-blood patients—an essential dimension of the clinical context (see Figure 1). As a developer working on illumination remarked: "For me, they're just images. I don't see the context. They're numbers, pixels. And I don't think there's anything wrong with that" [E40, 1:00:26]. Indeed, this detachment is not problematic in itself, as the removal of certain contextual elements enables the feasibility assessment to proceed and fosters deliberation about the system's applicability. Particularly, when considering the relationship between plates and lighting:

> We are more confident when it comes to canonical plates. For those with special traits, we deliberately request pictures with pathogens. This allows my colleagues to evaluate the best illumination settings, ensuring they include high contrast so that no growth is missed, whether very small or transparent. [E14, 18:50]

Consensual data collection represents a shift through which large volumes of digital images become incorporated into the AI infrastructure (Crawford, 2021). This also constitutes a key company asset for several reasons. First, the sheer volume of images downloaded, combined with the number of automation ecosystems deployed across the market, generates a resource that is difficult to replicate. Second, the analysis of these images for research and development purposes provides valuable insights into the diversity of laboratory practices. As one developer explained: "That is our strength. They [clients] know their own reality, limited to their lab, but we've gained knowledge over time about how various labs operate" [20.2, III, p. 39].

Importantly, the downloaded images are far from raw data. Despite a fundamental detachment, they retain elements of their clinical context in the form of metadata (labels) indicating the results (clicks) given by laboratory technicians when identifying the presence and type of bacterial growth. Thus, in developers' hands, these collected images function simultaneously as inputs and outputs, questions and answers, thereby constituting a valuable infrastructural resource. Taken together, consensual data collection emerges as a central practice of what Jaton (2021) terms 'databasing': the production of reliable data sets through which algorithms are trained and validated.

Seen in this way, the feasibility assessment foregrounds artificial vision as a central site of product customisation. It also opens an analytical space for examining practices of ground-truthing and datafication, through which digital images are progressively detached from their clinical context and transformed into a valuable AI infrastructure. It is within this experience of hesitation that AI outputs can be meaningfully compared with laboratory technicians' visual judgments. As one developer explained: "We run the software and check if the outputs match the ones they [technicians] have generated." [E13, 1:08:34]. These comparisons, in turn, inform decisions about whether existing algorithms can be efficiently applied or whether further development is required.

Image analysis software

The second technology is the image analysis software: a set of algorithms developed through machine learning techniques (see Figure 1, IAS). From a connectionist perspective, these algorithms represent the core of the IS's calculative capacity. Drawing on their account of the two main traditions in AI research—symbolic and connectionist—Cardon et al. (2018) argue that intelligent machines bring together three key elements: a world, a calculator, and a target. One influential strand of the connectionist approach, deep learning, relies on the possibilities afforded by an intensively datafied world, which is mobilised into a calculative space where inductive learning can occur. In this framework, learning emerges from large numbers of distributed calculations carried out across a hidden network of nodes (neurons) that process vast quantities of data by adapting their connections. The output of the calculator is thus largely the result of this connectionist, inductive learning, as designed and configured by AI experts.

When outputs are judged sufficiently accurate, the calculator is stabilised into software packages—commonly referred to as algorithms—that can be integrated into an AI-based product. Yet accuracy remains a provisional achievement. Even when algorithms are blackboxed within clinical IS applications, they persist as *algorithmic propositions* whose value is continually tested through ground-truthing practices (Jaton, 2017). In the case of plates with an opaque side, AI developers realised that no existing algorithms could handle the specific agar–colony relationship involved—namely, white and transparent microorganisms growing on opaque agar. This limitation triggered a period of innovation, during which a new neural network was developed. The result was a novel algorithmic proposition that was ultimately deemed valuable enough to be delivered to the client.

The resulting innovation can be understood as crystallising into a valuable intangible asset for the company. As one AI developer observed: "The IS can now recognise even the smallest colonies, including the transparent ones that are really hard to spot. This is all thanks to the development work we now have at our disposal." [20.1, II, pp. 32–33]. However, the attribution of value to new algorithms is not a unilateral decision—it requires client approval. For this purpose, AI developers issue a performance report in which the reliability of the IS is quantitatively evaluated in relation to laboratory technicians' assessments. Results are classified into four categories: (a) results in agreement; (b) discrepant results with low impact on productivity and safety (e.g. false positives versus false negatives); (c) discrepant results with medium impact if left unreviewed by technicians (e.g. affecting suggested analytical workups); and (d) false negative results produced by the IS, which may have a high impact if not reviewed.

On this basis, we stress that, in industrial contexts, close interaction with clients constitutes a rich source of incremental innovation, insofar as hesitation around connectionist options can gain momentum through such interactions. In the same vein, we note that the capacity to detect microbial growth and sort plate images depends on a shared consensus regarding optimal system performance, calibrated against the gold standard of laboratory technicians' judgment (Yrivarren, et al. forthcoming).

Expert system

The third technology is the expert system (see Figure 1, ES). Since their inception, symbolic machines have primarily sought to encode sets of if–then rules within computational frameworks that function as knowledge bases for deriving truth-like conclusions. These rules consist of symbols—conventional representations of a world that, within this paradigm, is conceived as external to the mind. Expert systems, a 1980s instantiation of symbolic machines, were designed to address real-world problems by modelling them through heuristics—that is, combinations of domain-specific concepts and logical operations elicited from experts (Brock, 2018; Cardon et al., 2018; Stevens, 2022). Biological work has been a particularly favoured testing ground for such systems—in our case, that of clinical microbiologists.

A sound understanding of laboratory protocols and heuristics is essential not only for assessing the performance of a specific system application, but also for the broader process of product customization. As a product advertisement states, "[IS] does not impose itself on the laboratory; rather, it automates the laboratory's rules." [3.1, I, p. 29]. By eliciting and embedding these rules, the IS differentiates itself by enabling clients to perceive the device as specifically tailored to their needs.

Nevertheless, as with any meaningful activity, the process of translating rules of thumb into formal decision trees proved challenging, particularly because such rules are deeply embedded in routine laboratory work. Their translation required the combined efforts of clinical microbiologists, AI developers, and the company's application specialists—the latter acting as intermediaries between these groups. This joint effort involved to carry out a further *detachment*, comparable to the downloading of plate images. As with the evaluations concerning illumination settings and algorithmic packages, the knowledge generated through the analysis of laboratory heuristics emerged as a valuable asset for the company. The commercial relevance of this detachment was highlighted by the imaging team manager:

> If we don't know when to set a threshold, we'd just be creating algorithms that have no real use. If I can't figure out how the algorithm will be applied, I can't put it on the market, and that means it won't be valuable to the user. [E13, 57:20]

In short, without formalised heuristics there can be no real-world application, and thus no use value can be attributed to the IS. It is therefore through the negotiation of classification rules that a core component of the market assemblage is placed at stake.

Orchestrating the encounter

In this section, we describe how AI developers orchestrated the negotiation concerning the worthiness of observing and handling a small number of controversial microorganisms. These infinitesimal entities became visible only when algorithms detected their presence in plate images, thereby calling into question the skilled eye of laboratory technicians. As we suggest, clarifying the rules to be embedded in the expert system was essential to completing the customisation of the IS and securing its adoption by the client.

Microcolonies as agents of discrepancy

During the performance evaluation, AI developers observed that the imaging algorithms detected microcolonies and subtle growth patterns, which the expert system then sorted into a dedicated folder for further review by laboratory technicians. By comparing the ground–truth results—those assigned by technicians during routine clinical work—with the outputs generated by the IS, the developers inferred that, in some cases, technicians had overlooked these entities and classified them as negatives. In other instances, technicians appeared to focus primarily on predominant growth while disregarding the presence of such minute colonies. As a result, *the IS proved capable of detecting features that even skilled human observation would neglect*, generating a discrepancy between the visual assessments of the algorithms and those of the technicians.

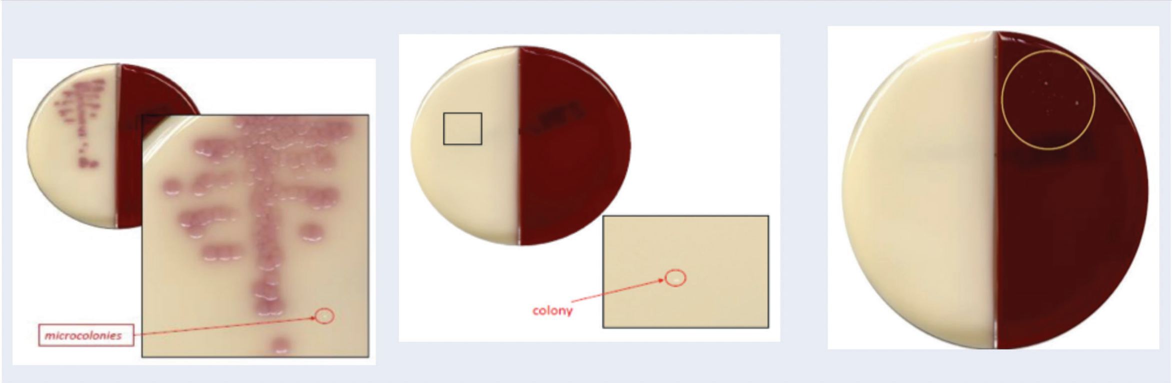
While the IS suggested retaining these entities that it had rendered visible, technicians considered them dispensable. In other words, from the perspective of the IS, infinitesimal forms of growth were worth noticing or, at the very least, warranted further examination by professional eyes. From the technicians' standpoint, however, such growths were either not visible during routine screening of plate images or, if observed, were not considered significant enough to justify specific follow-up analyses. At this point, two divergent answers emerged to the same question: *are microcolonies worth seeing in a clinical setting?*

The analytical framework presented so far—within which artificial vision, imaging algorithms, and the expert system mediate the relationship between the laboratory and the company—ultimately led to the socialisation of microcolonies as *agents of discrepancy*. From this perspective, hesitations surrounding the worthiness of seeing these tiny non-human entities triggered an exceptional situation: AI developers were required to engage directly with laboratory staff to negotiate their significance. As one developer remarked, "We have already achieved a certain level of knowledge that allows us to interact with the lab only in extraordinary cases." [E14a, 51:03]. Resolving this discrepancy thus became a necessary condition for finalising the customisation of the IS and, by extension, the commercial transaction.

AI developers framed the meeting by clearly outlining the discrepancy cases, the scope of the negotiation, its boundaries, and the potential courses of action. The discrepancy centred on exemplary cases grouped into three categories (Figure 2), consisting of plate images that had been detached from the performance assessment and recontextualised within a discursive presentation. Despite the developers' efforts to highlight the infinitesimal entities using arrows and circles, these entities remained extremely difficult to discern on the PowerPoint slides.

Similarly, the developers defined the scope of the negotiation by restricting it to reaching an agreement on the classification rules for the expert system. There was no open discussion of other aspects, such as the AI's visual capabilities. On the contrary, the developers sought to emphasise the effectiveness of the imaging algorithms. For example, a recurring message across several slides stated: "The AI highresolution software can detect microcolonies." In line with this framing, the developers proposed several alternative solutions: (a) setting specific thresholds to disregard microcolonies and classify plates as negative; (b) creating a new rule to categorise images into an appropriate folder (e.g., Not Significant Growth); or (c) validating the current IS results.

As Callon (1998, 1999, 2021) has argued, for a market transaction to take place, a space must be created in which heterogeneous evaluations can be weighted. In such a space actors should be allowed to define *multiple states of the world*, rank them according to their preferences, and assess alternative courses of action along with their potential consequences: "It is by allowing each agent to have preferences, to hierarchize them, and then to reveal and negotiate them—in a word, to calculate his or her interests, express them and defend them—that transactions are allowed to take place…" (Callon, 1998: 256). The orchestration of the encounter with laboratory members, aimed at weighting how best to integrate the IS into the laboratory's workflow, provides a clear illustration of this argument. However, enabling preferences concerning the inclusion or exclusion of microcolonies to be articulated proves to be a particularly challenging endeavour, as we show next.



Case 1. Presence of microcolonies

Case 2. Weak growth: white or transparent colonies

Case 3. Growth only on the red side of the plate

Figure 2. Minute colonies discussed.

Cognitivist approach to negotiation: The issue of domain

The main challenge lies in establishing a shared understanding of what constitutes the *domain* at stake. In symbolic AI, the term 'domain' has come to denote 'real world', although in practice it mainly refers to the process of translating situated, specialised knowledge into expert systems (Ribes et al., 2019). So, to grasp the expert knowledge properly is to delimit the real-world domain accurately. This notion encapsulates the principles of a cognitivist approach to mind and action that has been deeply embedded since the origins of AI in the 1950s, alongside a contemporaneous strand of social research concerned with cognition in non-Western cultures—commonly known as ethnoscience (Goodenough, 1957; Sturtevant, 1964; Werner, 1969). The common thread here is the assumption that, to understand what a domain is, a method is required for eliciting idiosyncratic classification categories, objects, and their relationships. It is further assumed that such a method makes it possible to render others' behaviours intelligible.

For this reason, although AI developers acknowledge that they are engaging with a world markedly different from their own, they cannot adopt a 'two cultures' thesis—as discussed in the introduction, which invokes the supposed divide between dry and wet modes of knowledge production, or between technocratic and care–oriented approaches to ethics. Adopting such a stance would imply that determining the IS's use value in clinical settings is unattainable. On the contrary, developers are keen to explore and translate local classification systems.

The cognitivist orientation of the negotiation is evident in developers' discourses, which emphasise the need to *overcome technicians' subjectivity* in order to establish a general framework for laboratory practices aligned with algorithmic capabilities.

> There aren't universal rules that labs follow in their routine work. Often, these rules are subjective, and different technicians apply different ones. The task is to develop a general rule that works for the entire lab and make sure it's compatible with our software's ability to generate accurate information. [E13, 1:06:32]

Consequently, when creating general rules in practice, developers and application specialists collaborate with microbiologists—typically senior supervisors or managers—whose perspectives are regarded as more consistent, knowledgeable, and authoritative. As a result, it is the microbiologists' judgement, rather than that of the technicians, that becomes the primary source of these rules—a pattern that is not unfamiliar in the history of science (Shapin, 1989):

> The results I download come from different lab technicians who rotate, so they're not always consistent with each other. […] But the requests I make are usually directed to either the head of the lab or the senior microbiologist. So, there's a bit of a disconnect, since technicians might not have the same knowledge as the microbiologists. The senior microbiologist, though, can give me a more objective answer. [E40, 19:25, 23:34]

The difficulties involved in establishing a domain also extend to language. Exploring heuristics forced laboratory members to resort to *metaphorical language* in order to express what they do every day—their lived experience—as linear actions following if–then rules. This effort can prove challenging even for experienced microbiologists. Thus, an AI developer acknowledged in front of laboratory members (as reported in the first excerpt below) that not all practices can be rendered in logical terms, echoing Collins's (1987) notion of the 'pain of regress' associated with the formalisation of cultural skills. This recognition, in turn, led developers to acknowledge limitations in the way IS's domain is constructed (as illustrated in the second excerpt):

> I know it's tough to put into words all the rules you use every day, and these meetings definitely bring out some unestablished rules. But the goal is to work together to resolve any discrepancies. [20.3, II, p. 44]
>
> We realise that we don't know a lot about the lab processes. There are so many complex rules of interpretation that some lab members suggest we should just stick to the traditional methods. [3.3, I, p. 36]

In addressing obstacles related to heuristic formalisation—namely, the selection of competent lab interlocutors and putting local rules of thumb into words—developers actively engage with an issue that has preoccupied STS since its inception: how authoritative knowledge is stabilised, and how practices—often chaotic—that contribute to its production are smoothed out. (Knor-rCetina, 1981; Latour and Woolgar, 1986). The cognitive approach sets a condition of possibility for negotiation: multiple states of the world can be discussed—that is, different ways of handling tiny microbial entities—provided that an authoritative domain is established in the process. Put differently, no use value can be attributed to the IS unless a version of real-world rules is encoded within its symbolic machine.

Contrasting visibility statuses of microcolonies

We now show that negotiating classification rules was not feasible without simultaneously settling the visibility status of microcolonies. Meetings between AI developers, the laboratory manager, and the laboratory engineer provided the setting for discussions about what is worth seeing, held at the height of the COVID19 pandemic.

In analysing the fieldwork material, we identified four distinct points of view concerning the visibility status of microcolonies (Table 1). Situated within different evaluative practices and scenarios, none of the actors—human or non–human—observe exactly the same thing. Rather, each actor, through a situated form of vision, enacts specific visibility statuses that produce distinct effects on plate images containing such infinitesimal forms of growth. These effects may be read as practical responses to the question that gives this work its title: *what is worth seeing?* Building on the discrepancy emerged between technicians' vision and that of the IS, we next describe the contrasting viewpoints of AI developers and senior laboratory members regarding how best to manage these entities. We also show that the negotiation process came to an end through modifications to the material–semiotic configuration of the system.

"Those unfortunate beasts"

AI developers take for granted that the IS possesses superior visual capabilities compared to humans. However, regardless of how accurately the algorithms detect minute colonies, developers recognise that these entities often do not exist in the eyes of laboratory technicians. In other words, under certain circumstances, technicians may be unable—or may choose not—to recognise them as valuable epistemic objects within the laboratory's analytical processes.

> …the lighting creates a strong contrast between the surface and the pathogens, letting the algorithms pick up even the tiniest details with amazing precision. But human operators don't see those tiny things, like microcolonies… For them, they just don't exist. […] We can detect almost

Table 1. Actors' situated viewpoints and the effects of different microcolony visibility statuses.

ACTORS	**Object of observation**	**SITUATED CONTEXT**	**Microcolonies' visibility status**	**Effects produced**
Lab technicians	Preprocessed plate images displayed on a web interface	Routine laboratory practices	Not seen	Microcolonies dismissed or subsumed
Imaging System	Datafied plate images, consensually downloaded	Ground-truthing and feasibility evaluations	Seen	Plates flagged for further review by technicians
AI developers	Plate images managed by the IS	Overall system performance evaluation	Seen	Microcolonies can be managed via rules or else dismissible
Senior lab members	Selected plate images presented in PowerPoint slides	Virtual meetings aimed at resolving discrepancies	Seen	Plate images contested and reversible.

> everything, and then it's up to us to decide whether to ignore it or not, based on what the lab actually does [E14s, 16:33]

A key part of AI developers' work consists in finding a *balance between what the AI system can detect and what a human technician can see*. In other words, their task is to determine whether paying attention to "beasts"—a term developers use to refer to microorganisms—is relevant within the context of specific laboratory routines. Some developers tend to regard "those unfortunate beasts" as negligible, primarily for two reasons. First, their experience of working with laboratories that process high volumes of samples through standardised procedures, in which large numbers of negative plates are routinely discarded each day. Second, the belief that the value of the IS lies in its capacity to optimise large-scale processes. This assessment of the worthiness of *not* seeing microcolonies is clearly illustrated in a conversation that took place shortly after a negotiation meeting:

> *Manager:* They're too theoretical. They don't see the practical side, like how the software detects things they can't. […] According to theory, whenever you observe colony growth—no matter how small—you're supposed to investigate it. But in practice, from my experience working with various labs, those insignificant microcolonies are either discarded or considered part of the same microorganism as the main growth. No one pays attention to those unfortunate beasts. In labs with large sample volumes, microcolonies are just thrown out because there's so much to process—unlike here. In this lab, the volume is much lower, and instead of dismissing microcolonies by set rules, the lab members are asking us to gather information about their processes. […]
> *Researcher:* In your view, what is the IS useful for?
> *Manager:* Generally speaking, the IS adds value to bulk processes—that is, labs with high volumes and standardised internal procedures. Put simply, it's useful because it sorts all the negative plates into one folder, letting you discard them with a single click. But for this lab, we need to think about how to adapt the system to their processes and volume to add value for them. [20.1, II, pp. 31-32]

Indeed, once it had been established that the algorithms could detect these infinitesimal entities—which might hold potential relevance for laboratory analysis—it became crucial to devise a strategy for adapting the IS's visual capabilities to the specific requirements of the client laboratory. This laboratory sought to enhance its operational capacity through technological innovation. Although it was not yet a high-throughput laboratory, it aspired to become one soon. In the laboratory manager's words:

> Our goal is to minimise the amount of work still being done manually and to increase our sample processing capacity. We believe the IS will be a great help with this, especially since we are a small lab that processes only a few samples per week and lacks standardised procedures. [20.3, II, p. 45]

Regardless of their accumulated experience, AI developers acknowledge that the fate of tiny microbial growths is context-dependent and shaped through processes of customisation. As Coopmans (2010) has noted, responsibility for judging the reliability of images typically rests with customers. In our case, the burden of assessing the value of observing microcolonies—whether as significant epistemic objects worthy of further investigation or as irrelevant "unfortunate beasts"—fell on laboratory members.

Certainly, this does not imply that clients are always right. On the contrary, as one developer noted, "Sometimes we suspect that [lab technicians] are making a mistake. So, we tend to push them into discussions to clarify their rules." [20.2, III, p. 39]. It can therefore be argued that, by opening a dedicated space for discussing the visibility statuses of microcolonies, AI developers *are, to some extent, shaping the clinical value of certain infinitesimal entities*, thereby prompting a revision of habitual laboratory's visual practices—and their effects on plate interpretation.

Image reversibility

Laboratory members were exposed to the outputs of a system they had not previously used, positioning them as 'pre-users'—that is, individuals who establish a relationship with medical devices that differs from the acceptance and adoption typically assumed to characterise use

(Kelly and Matthews, 2014). Their engagement with images and imaging was marked by contestation. In the excerpts that follow, the laboratory manager raises two concerns: first, the *representational accuracy* of the proxies she was observing, specifically several images displayed in a PowerPoint presentation; and second, the *retrospective method* used to decide among the proposals put forward by the AI developers. In doing so, the manager shifted her evaluation of worthiness back toward laboratory practices, prioritising tactile engagement over visual assessment.

> I [Field researcher] overheard the lab manager saying, "It's really hard to make decisions in hindsight," referring to the challenge of accepting or rejecting case-related proposals. "We can't approve a proposal based solely on a plate shown in a PowerPoint. Maybe the issue isn't with the rule, but with the incubation time." [20.1, II, p. 30]
> "We don't need more examples; we want to have the plates in our hands." [20.1, II, p. 31]

Laboratory members insisted on recontextualising the proxies by *turning them upside down* and reintegrating physical plates and plate images into their analytical practices, even at the risk of exacerbating discrepancies and tensions in the negotiation. Unsurprisingly, this approach clashed with the developers' cognitivist orientation. For example, the laboratory manager emphasised the circumstantial nature of microcolony analysis: "In cases where we have normal urine samples, without clear symptomatic infections, it might be an option to consider microcolonies < 10,000 [$< 10^4$ CFU] as negative. But we need more cases before deciding." [21.1, I, p. 47]. This "but" is key here, as the manager was thereby suggesting that ambiguous or atypical cases are a recurring feature of clinical knowledge production, rendering any attempt to formalise them vulnerable to the *pain of regress*.

The encounters grew increasingly complex as laboratory members introduced *unexpected issues* that extended beyond the original scope of the negotiation. This became particularly evident when the laboratory engineer asked for a precise definition of what constitutes a microcolony:

> After a brief exchange, during which the imaging team manager loosely suggested some elements of the definition, a young developer hesitantly said, "A microcolony is a small colony…" This prompted laughter from the laboratory. "It's important for us to know what criteria you're using," the lab engineer interjected. Then, with new-found confidence, the young developer stated that a microcolony is defined by what the AI software identifies based on size (about 0.1 mm in diameter) and colour (either one that doesn't match the target colour or one with no colour). [20.1, II, pp. 32, 36]

Requesting the definitional criteria used to develop an algorithmic proposition reflects an intent akin to a cognitivist approach to the formalisation of rules. Each party was keen to understand how the other defined the aspects it considered relevant. This *shared pain of regress* becomes evident in the young developer's response. Although exploring definitions may appear theoretical, it has a distinctly practical dimension. It can be interpreted as an attempt by laboratory members to assess the AI developers' domain expertise and, by extension, to judge the reliability of the technology. Moreover, probing definitional criteria places developers under scrutiny and fosters greater awareness of the *performative nature of imaging devices*: a microcolony is defined by what the AI identifies as such after applying specific calculative criteria.

The negotiation became even more pressing when addressing type 3 cases (Figure 2), which exhibited growth exclusively on the plate's red side. In analysing these images, the IS applied different algorithms to each side of the plate. On the left side, algorithms detected and counted overall bacterial growth (colonies) and classified it by morphology, thereby identifying isolate predominance and quantity. On the right side, distinct algorithms were applied solely to detect and count colony growth. Unlike the other cases discussed, the IS lacked default rules within its decision tree for type 3 cases. As a result, it was unable to propose classification outcomes for these plates. What, then, should be done with this infinitesimal growth? Was it worth seeing and how?

From the AI developers' viewpoint, growth on the right (red) side could be considered relevant only if it coincided with predominant growth on the left (white) side. Conversely, if growth was observed solely on the right, it was largely deemed meaningless and could be disregarded. In contrast to this evaluation—and far from yielding a clear-cut rule that could be incorporated into the expert system—laboratory members raised strong *objections to the configuration of the algorithms used to analyse the images*. Crossing previously established negotiation boundaries, the laboratory engineer asserted that the algorithms should analyse the plates in the same way they themselves had been doing:

> All growth needs to be assessed by considering both sides. Otherwise, how can you tell if there's mixed growth? We evaluate both sides using the same criteria. And it's not accurate to say that growth can't be verified on just the right side. In those cases, just like with the left side, you have to apply the same parameters for colour, microbe class, and load. [20.3, II, p. 43]

All things considered, we coin the term 'image reversibility' to describe the process through which laboratory members evaluate the worthiness of seeing microcolonies. Rather than directly translating their expertise into rules that would expand the AI's knowledge base, these pre-users sought to understand how the IS was able to detect such entities. In doing so, they encouraged AI developers to disclose certain aspects of the algorithms' design. The concept of image reversibility thus helps trace the backward movement of images to their conditions of production, echoing Latour's use of the rhetorical notion of 'modalities' (Latour, 1987, 1999; Latour and Fabbri, 2000) to highlight variations in the truth value of statements: "modalities […] often acted like price tags of statements, or, to use a mechanical analogy, like an expression of the weight of a statement." (Latour, 1987: 84). Similarly, rendering an image reversible functions here as a *contestation tag*, which unfolds as follows

1. Dislocating the evaluation of worthiness from a set of disentangled proxies (or discrete pieces of information) and relocating it within local cognitive practices that maximise sensory engagement with specimens, tools, and living entities.
2. Introducing unforeseen matters of concern into deliberation to expand its scope, by examining the definitional criteria embedded in machine learning and revealing how algorithms participate in shaping reality.
3. Challenging the counterparty's evaluations of worthiness by pushing the boundaries of discussion and disputing how algorithms should interpret the world.

Image reversibility thus serves as a means of problematising the conventions that underpin the implementation of an AI-based medical imaging device.

Transaction unlocked

Following the negotiation meetings, both parties continued to refine an agreement aimed at finalising the implementation. One year later, the laboratory accepted the initial proposal to replace unknown opaque plates with known clear ones, likely because, as we were informed, the laboratory leadership team involved in the IS implementation had been removed and replaced. This decision prompted a sudden shift from negotiation to the material–semiotic conditions that enable both non–human entities to grow and artefacts to render microorganisms observable.

Building on this new materiality, a different illumination setting was introduced, an alternative algorithmic proposition was developed, and a key rule was formalised. Regarding the latter, the initial performance assessment specified among the imaging criteria that microcolonies should be disregarded if they numbered fewer than ten colonies (10^4 CFU) *in the presence of* other predominant isolates; otherwise, these minute entities were to be considered an isolated group. This formulation was revised in the final assessment, which established that, to discard these entities, their count must be less than or equal to nine (10^3–10^4 CFU). *However*, if a predominant bacterial isolate (e.g. Enterococcus, Streptococcus) is detected simultaneously, the microcolonies must be taken into account and the plate sent for further review.

The primary difference between the two statements lies in how the rule was refined to align with the laboratory's requirements. The refinement increased the likelihood that these entities—whether present alone or alongside predominant growth—would be incorporated into laboratory knowledge production.

Taken together, these subtle modifications to the material–semiotic configuration of the system—jointly enacted by both parties through long–term interactions—endowed microcolonies with the status of worthy epistemic objects, to be seen according to agreed-upon criteria. Only then was the commercial transaction unlocked.

Final remarks — on 'artisanal intelligence'

Before an IS, as a finalised product, can be transferred and put into use, its adequacy for the laboratory's workflow must be validated by those who will ultimately be held accountable for the system's performance. In other words, its use value must be confirmed by laboratory members—namely, senior microbiologists and managers. One might therefore argue that what occurs through negotiation is the social shaping of certain technical features of the system so as to fit existing laboratory's interpretative practices. However, as we have shown, detecting minuscule entities—barely visible and difficult to classify—induces a state of hesitation that extends across a series of evaluations concerning the system feasibility, accuracy, and performance during the implementation process. It is this state of hesitation that drives the orchestration of a dedicated (and extraordinary) negotiation with clients. Viewed from this perspective, neither laboratory's visual practices nor imaging capabilities can be taken for granted without eroding the world-making implications of negotiation.

We therefore argue that evaluations and negotiations constructively destabilise both system capabilities and entrenched laboratory's ways of seeing, giving rise to a *process of co-producing a socio-natural reality*. Accordingly, several questions emerge and become entangled in this process: How can idiosyncratic heuristics be formalised? Who counts as a legitimate interlocutor with domain expertise? How can the transaction be steered towards a successful outcome? Is it worth showing small or transparent growths? How should these entities be handled? Under what conditions do microcolonies come to deserve the status of epistemic objects? The epitome of this co-production resides in the very question the actors are jointly striving to answer: *What is worth seeing?*

What the negotiation analysed here reveals is that there are no visual practices that the IS can automate without significant investments of time and money to shape them, alongside adjustments to its own technological configurations. Conversely, no client laboratory can shape the IS without comparable investments to ensure that their requests for changes in imaging capabilities and rules encoding are accommodated—while simultaneously accepting that their tacit interpretative practices be put up for discussion. This *symmetrical realisation* allows us to grasp the sharp (dis)connection between those who enact a 'cognitivist approach' to negotiation and those who resist it through what we term 'image reversibility'. It is this friction that emerges as the leitmotif of the co-production process, and more concretely, of the exchanges among the actors concerned.

That said, how should we name this form of co-production to make its specific characteristics visible? We encountered the term 'artisanal intelligence' incidentally during a discussion with the company's CEO about an early draft of this paper. On that occasion, he remarked: "Mostly, what we do is not AI, in capital letters, but artisanal intelligence. We tailor it to fit." 5 In doing so, he articulated the organisation's distinctive approach to AI implementation—namely, its capacity to adapt to different local contexts (i.e. laboratory routines) by embedding those contexts into the system's configuration. More broadly, we suggest that artisanal intelligence may serve as a compelling sociological *metaphor for a mode of AI world–making* that relies on ongoing hesitation, complex negotiations, and ad hoc tailoring. In conclusion, we invite reflection on the implications of this notion in light of current STS debates.

First, with regard to technological development, the artisanal style challenges the assumption that connectionist and symbolic approaches in AI are inherently antagonistic, regardless of their context of application. Within the IS architecture, these approaches are *not only complementary but also subject to evaluation*. Moreover, without integrating artificial vision, deep-learning algorithms, and expert systems in a way that is accurate, consistent and consensual, the company would struggle to address clients' specific needs. In this respect, although we build on their conceptual framework, we depart slightly from Cardon et al. (2018), whose primary focus is on tracing the diverging trajectories of intelligent machines—that is, their distinct world–calculator–target configurations. Our focus, by contrast, has been on the specificities of implementing an AI-based medical device: one that integrates multiple machines and requires diverse forms of interaction between developers and customers.

Second, from an ontological stance (Law and Lien, 2013; Lynch, 2013), we assert that *the IS's true target is the world itself*—that is, the reshaping of key laboratory practices and epistemic objects (i.e., specimens and living entities) emerges as a necessary goal in order to demonstrate both feasibility and value. This world-making trait provides the calculators—being connectionists and/or symbolic—with its working domain. As we have observed, the IS's world must be constructed prior to its actual implementation: before the IS could process inputs, the client laboratories must acquire the company's automation ecosystem, which enables the datafication of culture-plate images through an artificial-vision infrastructure (referred to as *Detachment* in Fig. 1).

Likewise, once client laboratories have validated the IS performance—either after negotiation, as we have shown, or by unquestioned acceptance—the world in which they work on a daily basis is no longer the same, at least in two respects: (a) it has been inscribed into a series of connectionist–symbolic calculators (AV, IAS, and ES in Figure 1), and (b) it has been reshaped to allow the automatic classification of digital plates based on presumptive pathogen identification (*Reattachment* in Figure 1). Importantly, this world-making step prompted by IS implementation unfolds through that uncertain, negotiated, and tailored-to-fit co-production that constitutes the core challenge of artisanal intelligence.

Third, from a product-management perspective, we suggest that the key distinction in customisation lies between merely reproducing conventional know-how and embracing hesitation as a productive force. Indeed, the pragmatist conception of intelligence is grounded precisely in this distinction (Dewey, 1930). AI developers undoubtedly know how to implement an IS; they possess the requisite expertise. However, throughout the process, they encounter novel challenges that put both the system's capabilities and their own knowledge to the test. Jaton (2021) has posed a provocative question in this regard: which circumstances foster the exploration of hesitation? From our perspective, *client involvement in product customisation is central to this dynamic, as it introduces unforeseen requests*. In an industrial setting, AI developers may experience hesitation around genuine development options precisely through listening to and negotiating with clients. The key challenge for product-management practitioners, then, lies in how to collect, interpret, and derive value from these heterogeneous requirements—without neutralising productive hesitation through excessive standardisation or bureaucracy.

In synthesis, we propose 'artisanal intelligence' as a notion to describe, in a symmetrical manner, the co-production of socio–natural realities prompted by the implementation of AI-based imaging devices in clinical laboratories. Crucially, this notion foregrounds the ways in which actors involved in market transactions navigate a paradox inherent to these devices: while their calculative power enables the detection of an abundance of visual details, this very capability can undermine their use value in real-world settings. One key method in which this paradox is managed is through the negotiation of what is worth seeing.

Acknowledgements

We extend our sincere gratitude to Mario S., CEO of the company, for granting the field researcher full autonomy in conducting the study and for providing valuable feedback. We also thank all members of the R&D department for their contributions, with special thanks to Andrea B., Nicolò C., Guido S., Stefano O., and Arnalda G. Joaquín Yrivarren expresses his heartfelt thanks to Jaime O. Salomón for his trust and support.

Joaquín Yrivarren discloses that, after completing the fieldwork, he joined the company's Project Management Team and is currently employed as a project coordinator in the R&D Automation Department at the time of this paper's publication. Miquel Domènech declares no conflicts of interest.

References

Aeffner F, Zarella M D, Buchbinder N et al. (2019) Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. *Journal of pathology informatics* 10(9).

Argudo-Portal V and Domènech M (2022) 'We came here to stay': Making biobanks worth maintaining in Spain. *Sociology of Health & Illness* 44(2): 328–344.

Becker H (2011) *Trucos del oficio: cómo conducir su investigación en ciencias sociales*. Buenos Aires: Siglo Veintiuno Editores.

Brock DC (2018) Learning from artificial intelligence's previous awakenings: The history of expert systems. *The AI Magazine*, 39(3): 3–15.

Callon M (1986) Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. In: Law J (ed) *Power, Action, and Belief: A New Sociology of Knowledge?* London: Routledge & Kegan Paul, pp. 196–233.

Callon M (1993) Society in the making: The study of technology as a tool for sociological analysis. In: Bijker W E, Hughes T P and Pinch T J (eds) *The social construction of technological systems*. London: MIT Press, pp. 83–106.

Callon M (1998) *The laws of the markets*. Oxford: Blackwell Publishers/Sociological Review.

Callon M (1999) Actor-Network Theory—The market test. *The Sociological Review* 47(1_suppl.): 181–195.

Callon M (2021) *Markets in the making: Rethinking competition, goods, and innovation*. New York: Zone Books.

Callon M and Muniesa F (2005) Economic markets as calculative collective devices. *Organization Studies* 26(8): 1229-1250.

Cardon D, Cointet JP and Mazieres A (2018) Neurons spike back: The invention of inductive machines and the artificial intelligence controversy. *Réseaux* 211(5): 173–220.

Cochoy F (2008) Calculation, qualculation, calqulation: Shopping cart arithmetic, equipped cognition, and the clustered consumer. *Marketing Theory* 8(1): 15–44.

Collins, H M (1987) Expert systems and the science of knowledge. In: Bijker W E, Hughes T P and Pinch T J (eds) *New directions in the sociology of technology*. Cambridge: MIT Press, pp. 321–341.

Coopmans C (2010) 'Face value': New medical imaging software in commercial view. *Social Studies of Science*, 4(2): 155-176.

Crawford K (2021) *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press.

Dewey J (1930) *Human nature and conduct: An introduction to social psychology*. New York: The Modern Library.

Drogt J and Milota M (2021) The (in)invisibility of human tissue (journey of a dermatological mole). *The Polyphony*. Available at: https://thepolyphony.org/2021/06/10/the-invisibility-of-human-tissue-journey-of-a-determatological-mole/ (accessed 27 September 2024).

Drogt J, Milota M, Vos S, Bredenoord A and Jongsma K (2022) Integrating artificial intelligence in pathology: A qualitative interview study of users' experiences and expectations. *Modern Pathology* 35(11): 1540–1550.

Food and Drug Administration [FDA] (2021) *Good Machine Learning Practice for Medical Device Development: Guiding Principles*. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles (accessed 27 September 2024).

Fujimura J (1992) Crafting science: standardized packages, boundary objects, and translation. In: Pickering A (ed) *Science as practice and culture*. Chicago: University of Chicago Press, pp. 168–211.

Gad C and Ribes D (2014) The conceptual and the empirical in science and technology studies. *Science, Technology, & Human Values* 39(2): 183–191.

Goodenough W H (1957) Cultural anthropology and linguistics. In: Garvin P (ed) *Seventh annual roundtable of linguistics and language studies*. Washington: Georgetown University Press, pp. 167–173.

Harris C V (2011) Technology and transparency as realist narrative. *Science, Technology, & Human Values* 36 (1): 82–107.

James W (1907) La concepción pragmatista de la verdad. In: *Pragmatismo: un nuevo nombre para viejas formas de pensar*. Madrid: Alianza Editorial, pp. 149–168.

Jaton F (2017) We get the algorithms of our ground truths. *Social Studies of Science* 47(6): 811–840.

Jaton F (2021) Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application. *Big Data & Society* 8(1): 1–15.

Joyce K (2005) Appealing images: Magnetic resonance imaging and the production of authoritative knowledge. *Social Studies of Science* 35(3): 437–462.

Joyce K (2006) From numbers to pictures: The development of magnetic resonance imaging and the visual turn in medicine. *Science as Culture* 15(1): 1–22.

Kelly J and Matthews B (2014) Displacing use: Exploring alternative relationships in a human-centered design process. *Design Studies* 35: 353–373.

Knorr-Cetina K (1981) *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford and New York: Pergamon Press.

Lalumera E and Fanti S (2020) The philosophy of advanced medical imaging: Mapping the field. In: *Philosophy of advanced medical imaging*. Switzerland: Springer, pp. 1–9.

Latour B (1986) Visualisation and cognition: Thinking with eyes and hands. In: Lynch M and Woolgar S (eds) *Representation in scientific activity*. Cambridge: MIT Press, pp. 19–68.

Latour B (1987) *Science in action: How to follow scientists and engineers through society.* Cambridge: Harvard University Press.

Latour B (1998) The politics of explanation: An alternative. In: Woolgar S (ed) *Knowledge and reflexivity: new frontiers in the sociology of knowledge*. London: SAGE, pp. 155–176.

Latour B (1999) *Pandora's Hope: Essays on the reality of science studies.* Cambridge: Harvard University Press.

Latour B (2005) *Reassembling the social: An introduction to actor-network-theory*. Oxford: Clarendon.

Latour B and Fabbri P (2000) The rhetoric of science: Authority and duty in an article from the exact sciences. *Technostyle* 16(1): 115-134.

Latour B and Lépinay V A (2009) *The science of passionate interests: an introduction to Gabriel Tarde's economic anthropology*. Chicago: Prickly Paradigm Press.

Latour B and Woolgar S (1986) *Laboratory life: The construction of scientific facts*. Princeton: Princeton University Press.

Law J (2009) Actor-network theory and material semiotics. In: Turner B (ed) *The New Blackwell companion to social theory*. Oxford: Blackwell, pp. 141–158.

Law J and Lien M (2013) Slippery: Field notes in empirical ontology. *Social Studies of Science* 43(3): 363–378.

Lynch M (1988) The externalized retina: Selection and mathematization in the visual documentation of objects in the life sciences. *Human Studies* 11(2/3): 201–234.

Lynch M (2013) Ontography: Investigating the production of things, deflating ontology. *Social Studies of Science* 43(3): 444–462.

Muniesa F, Millo Y and Callon M (2007) An introduction to market devices. *The Sociological Review* 55(s2): 1–12.

Paranjape K, Schinkel M, Hammer R D et al. (2021) The value of artificial intelligence in laboratory medicine. *American Journal of Clinical Pathology* 155(6): 823–831.

Peiffer-Smadja N, Dellière S, Rodriguez C et al. (2020) Machine learning in the clinical microbiology laboratory: Has the time come for routine practice? *Clinical Microbiology and Infection*, 26: 1300–1309.

Ponterotto J G (2006) Brief note on the origins, evolution, and meaning of the qualitative research concept 'thick description.' *The Qualitative Report* 11(3): 538–549.

Prasad A (2005) Making images/making bodies: Visibilizing and disciplining through magnetic resonance imaging (MRI). *Science, Technology, & Human Values* 30(2): 291–316.

Ribes D, Hoffman A, Slota S and Bowker G (2019) The logic of domains. *Social Studies of Science* 49(3): 281–309.

Scherz P (2019) The displacement of human judgment in science: The problems of biomedical research in an age of big data. *Social Research* 86(4): 957–976.

Schiølin K (2020) Revolutionary dreams: Future essentialism and the sociotechnical imaginary of the fourth industrial revolution in Denmark. *Social Studies of Science* 50(4): 542–566.

Shapin S (1989) The invisible technician. *American Scientist* 77(6): 554–563.

Silvast A and Virtanen M (2023) On theory-methods packages in Science and Technology Studies. *Science, Technology, & Human Values* 48(1): 167-189.

Sismondo S (2010) *An introduction to Science and Technology Studies*. Chichester, UK: Wiley-Blackwell.

Stevens H (2022) The business machine in biology—The commercialization of AI in the life sciences. *IEEE Annals of the History of Computing* 44(1): 8–19.

Sturtevant W C (1964) Studies in ethnoscience. *American Anthropologist* 66(3): 99–131.

VallèsPeris N and Domènech M (2020) Roboticists' imaginaries of robots for care: The radical imaginary as a tool for an ethical discussion. *Engineering Studies* 12(3): 157–176.

VallèsPeris N and Domènech M (2021) Caring in the inbetween: A proposal to introduce responsible AI and robotics to healthcare. *AI & Society* 38(4): 1685–1695.

Van Baalen S J, Carusi A, Sabroe I and Kiely D G (2017) A social-technological epistemology of clinical decision-making as mediated by imaging. *Journal of Evaluation in Clinical Practice* 23(5): 949–958.

Werner O (1969) The basic assumptions of ethnoscience. *Semiotica* 1(3): 329-338.

World Economic Forum [WEF] (2019) *Health and health in the fourth industrial revolution: Global Future Council on the Future of Health and Healthcare 2016-2018*. Available at: https://www3.weforum.org/docs/WEF__Shaping_the_Future_of_Health_Council_Report.pdf.

Yoxen E (1987) Seeing with sound: A study of the development of medical images. In Bijker W E, Hughes T P and Pinch T (eds) *New directions in the sociology of technology*. Cambridge: MIT Press, pp. 273–297.

Yrivarren J, Jacomuzzi A C, and Domènech M (forthcoming) Promises / manual labor paradox: The accountability style of on-site validation in microbiology laboratory automation. *Science, Technology, & Human Values*. Accepted for publication.

Notes

1 In structuring our descriptions, we have adopted some framing practices proposed by Callon (2021) to explain how commercial transactions come about. These include the customisation of a transferable good, the orchestration of encounters to foster product-customer attachment, and the deployment of actors' evaluations. Notably, none of these elements—whether goods, encounters, or evaluations—are set in stone or confined to rigid compartments. Instead, they are intimately intertwined and shaped by unexpected events that prompt technical, cognitive, and social adjustments. Accordingly, we reject the notion of providence—whether in the form of an invisible hand or state intervention—as a force harmonizing economic exchanges from the outset (Latour and Lépinay, 2009). Rather, market transactions require continuous framing and reframing through a range of mundane and sophisticated tools that enable actors to make decisions (Callon, 1998, 1999; Cochoy, 2008; Muniesa et al., 2007). On this basis, we adopt Callon & Muniesa's (2005: 1245) understanding of a market as "a collective device for the evaluation of goods."

2 To describe the laboratory's perspective, we drew on interventions from online meetings and grey documentation.

3 This coding applies to the interviews conducted. It indicates the interview number and the time in the recording at which the quoted excerpt begins.

4 This coding refers to the organization of our ethnographic logbook. It specifies the week and day of observation, the entry number, and the page number.

5 Importantly, this statement emerged during feedback sessions on the first full draft of our manuscript, rather than through participant observation. While members of the imaging team recognise adaptability as a key feature of the IS, we do not claim that the term artisanal intelligence constitutes an established part of their everyday vocabulary.