

# Embedding artificial intelligence in healthcare: An ethnographic exploration of an AI-based mHealth app through the lens of legitimacy

DIGITAL HEALTH  
Volume 10: 1–20  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076241292390  
journals.sagepub.com/home/dhj



Sydney Howe<sup>1</sup> , Anna Smak Gregoor<sup>2</sup> , Carin Uyl-de Groot<sup>1</sup>,  
Marlies Wakkee<sup>2</sup>, Tamar Nijsten<sup>2</sup> and Rik Wehrens<sup>1</sup>

## Abstract

**Purpose:** Skin cancer, a significant global health problem, imposes financial and workload burdens on the Dutch healthcare system. Artificial intelligence (AI) for diagnostic augmentation has gained momentum in dermatology, but despite significant research on adoption, acceptance, and implementation, we lack a holistic understanding of why technologies (do not) become embedded in the healthcare system. This study utilizes the concept of legitimacy, omnipresent but underexplored in health technology studies, to examine assumptions guiding the integration of an AI mHealth app for skin lesion cancer risk assessment in the Dutch healthcare system.

**Methods:** We conducted a 3-year ethnographic case study, using participant observation, interviews, focus groups, and document analysis to explore app integration within the Dutch healthcare system. Participants included doctors, policy-makers, app users, developers, insurers, and researchers. Our analysis focused on moments of legitimacy breakdown, contrasting company narratives and healthcare-based assumptions with practices and affectively-charged experiences of professionals and app users.

**Results:** Three major kinds of legitimacy breakdowns impacted the embedding of this app: first, lack of institutional legitimacy led to informal workarounds by the company at the institutional level; second, quantification privilege impacted app influence in interactions with doctors; and third, interactive limits between users and the app contradicted expectations around ease of use and work burden alleviation.

**Conclusions:** Our results demonstrate that legitimacy is a useful lens for understanding the embedding of health technology while taking into account institutional complexity. A legitimacy lens is helpful for decision-makers and researchers because it can clarify and anticipate pain points for the integration of AI into healthcare systems.

## Keywords

AI, skin cancer, mHealth, legitimacy, health technology adoption, health system infrastructure

Submission date: 28 June 2024; Acceptance date: 27 September 2024

## Introduction

The potential of new, data-driven artificial intelligence (AI) applications in healthcare has been widely recognized.<sup>1–4</sup> AI-based healthcare applications are being introduced across a variety of medical disciplines, gaining momentum as augmenters of human-based diagnostics.<sup>5–8</sup> AI tools have proven especially promising for image recognition and analysis; clinically-validated diagnostic support tools

<sup>1</sup>Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

<sup>2</sup>Department of Dermatology, Erasmus Medical Center, Rotterdam, The Netherlands

### Corresponding author:

Sydney Howe, Erasmus School of Health Policy and Management, Erasmus University Rotterdam, PO Box 1738, 3000 DR, Rotterdam, The Netherlands..  
Email: howe@eshpm.eur.nl



are now available in disciplines ranging from radiology<sup>1</sup> to dermatology.<sup>5</sup> However, despite strong desire from developers, doctors, policymakers, and patients, many technically-sound and clinically-promising technologies struggle to become structurally embedded in healthcare.<sup>9–11</sup> Simultaneously, problems caused by flawed technologies that have become embedded have significant financial, medical, and social costs,<sup>12,13</sup> especially in terms of public trust.<sup>14,15</sup>

Dermatology is considered an especially interesting discipline for AI applications performing image classification due to the large number of clinical image databases available and the visual component of diagnosis.<sup>16,17</sup> In an effort to provide the right care at the right time and alleviate healthcare budget pressures,<sup>7,18</sup> mobile health (mHealth) applications using artificial intelligence (AI) to facilitate early diagnosis of skin cancer are in particular gaining popularity. Studies of skin cancer risk assessment apps often present potential benefits of these technologies in terms of improved accuracy of detection, better care access for patients, and work burden relief for physicians.<sup>5,6,19,20</sup> Yet despite widely recognized potential benefits, it is currently unclear how these technologies may reshape interactions among patients, doctors, and other actors within a healthcare system. Furthermore, it is not yet known how these contextual interactions may impact the adoption of the technology in the wider healthcare system. To our knowledge, this study represents the first ethnographic exploration of the adoption, acceptance, and implementation of a skin cancer risk assessment app, and one of the first studies of its kind for any AI healthcare technology.

In explorations of healthcare technologies, existing frameworks for studying technology acceptance, adoption, implementation, and governance provide significant data about parts of the process through which technologies become embedded in healthcare.<sup>21–29</sup> For instance, the Technology Acceptance Model focuses on acceptance via behavioral intention,<sup>26</sup> while other frameworks attempt to cover more issues at once using a variety of qualitative methods, such as the non-adoption, abandonment, scale-up, spread, and sustainability framework (NASSS).<sup>9</sup> We use the term “embedding,” in the sense of making a process (such as using technology) a normal part of everyday life (in a healthcare context).<sup>30</sup> Thus, we add to these frameworks’ understanding of individual processes work together to incorporate technology into healthcare in a way that eventually becomes taken for granted. Even with so much data, researchers, policymakers, and developers still have difficulties determining why some technologies become structurally embedded in healthcare while others do not. For instance, many technology acceptance and adoption studies are based primarily on studies of behavioral intention and other perceptions of the future among participants,<sup>26,28,31,32</sup> limiting their predictive power for context-based obstacles that are only made visible in practice.<sup>26</sup> These issues are well known and recently, there has been progress in

developing frameworks to account for these gaps, for instance, by focusing on organizational barriers<sup>33</sup> and outcomes not driven by acceptability.<sup>9,34</sup>

However, acceptance, adoption, and implementation frameworks rarely address legitimacy as a primary issue for embedding new technologies, despite the ubiquitous use of the term throughout the literature on health technology across several social science and medicine disciplines.<sup>35–40</sup> We argue that legitimacy is a crucial theme in understanding why and how technologies are (not) structurally embedded in healthcare.

We build on Suchman’s 1995 seminal definition of legitimacy as “a generalized assumption that an entity and/or its actions are desirable, proper, or appropriate within some socially-constructed system of norms, values, and beliefs” (574).<sup>41–44</sup> Drawing from literature within Organization & Management Studies and Science & Technology Studies, we add that legitimacy has an affective quality, in that it may contain sensations and feelings but it is not in itself an emotional quality.<sup>43</sup> We also add that legitimacy is always relational, in that it occurs between at least two entities (such as actors, institutions, or technologies).<sup>42</sup> While legitimacy studies related to healthcare and technology have explored the term in depth conceptually,<sup>45,46</sup> few studies have focused explicitly on legitimacy as a means of understanding how and why a new healthcare innovation or health technology becomes (or does not become) embedded.<sup>37,47</sup> Moreover, many studies view legitimacy of new technologies only in terms of language or narratives. Limited literature examines how legitimacy is produced through the practical actions and negotiations of actors within a healthcare system, despite empirical evidence of practice as a key driver of legitimation for new technologies.<sup>38,39,48</sup> This study does both: here, we focus on legitimacy as essential for adoption, acceptance, and implementation of health technologies through an ethnographic analysis that explicitly centres practices as a primary driver of legitimation. We use both narratives and relationships as a context for understanding expectations<sup>49</sup> that may help drive assumptions leading to legitimacy misalignments.

We take as our case study an AI-based skin cancer risk assessment smartphone app. The SkinVision app uses a convolutional neural network to produce a high- or low-risk rating for users based on images taken with a smartphone camera; photos are also checked by a team of dermatologists. It is trained using thousands of photos for which diagnoses have been confirmed by a dermatologist or histopathology when available. As of May 2024, the app is reimbursed by several insurance companies within the basic insurance package, providing over 8 million people<sup>50–53</sup> in the Netherlands with access to the app, even though it does not yet have approval from the Dutch National Health Care Institute (ZIN). Both this app<sup>a</sup> and others have been explored extensively in clinical trials;

however, we have found no ethnographic studies of how these apps are integrated into healthcare systems. Using legitimacy as our theoretical lens also answers calls for ethnographic research of AI tools that address ambiguous, subtle, and absent data adequately.<sup>54</sup>

Studying the role of legitimacy in embedding AI-based technologies in medical practice empirically can be challenging because elements within a fully-legitimate system can become so taken for granted they are rendered nearly invisible.<sup>55,56</sup> A focus on moments of legitimacy breakdown makes it possible to study assumptions and the backgrounded infrastructure<sup>55</sup> supporting legitimacy processes in a complex system, such as technology integration in healthcare. Assumptions and backgrounded infrastructure, in the context of this study, can include a wide variety of unquestioned and unnoticed elements in both the narrative around technology and the medical system. For instance, as other authors have noted, assumptions about health technology usage are built into technology designs,<sup>57–59</sup> and unspoken assumptions about hierarchies in medicine can make it difficult to shift practices even after change is implemented.<sup>38,60</sup> We include both assumptions built into the company narrative and assumptions ingrained within the medical context into which the app is embedded. We consider “breakdowns” in legitimacy to be moments when narrative or context-based assumptions are brought into question or actively contradicted through practices pertaining to the app or users’ affective relationship to the app. The existence of moments of legitimacy breakdown does not mean narratives are incorrect, or differences with practices are (necessarily) problematic. Rather, by focusing on these moments, we gain a deeper understanding of the nuance and complexity involved in embedding technology in healthcare because it allows us to observe and thus question baseline assumptions that otherwise go unnoticed during the embedding process.

Therefore, our research question is: what can we learn from moments of legitimacy breakdown about the assumptions guiding the embedding of this AI-based application in the Dutch healthcare system? To answer our research question, we combine a theoretical lens grounded in a long tradition of legitimacy studies in the social sciences while employing an ethnographic approach to study legitimacy empirically.

We first discuss our study design and methodologies used. We then explore our data with an analysis of the company narrative, followed by a discussion of how breakdowns manifest between this narrative and the practices of institutions (such as bureaucratic processes), users, and doctors. Next, we examine how the affective aspects of relationships between individual users and the app impact legitimacy. Finally, we discuss the implications of our data for technology acceptance, adoption, and implementation through the lens of legitimacy.

## Methods

This study was conducted using a wide range of ethnographic methods between 2020 and 2023. These methods included participant observation at general practitioner (GP) practices, dermatology clinics, work places, and personal homes; narrative interviews and open-ended focus groups with expert stakeholders; autoethnographic observations in fieldnotes; remote ethnographic fieldnotes made using digital tools such as Microsoft Teams and video recordings; and document analysis of research and strategy documents from Erasmus Medical Center (EMC) and the Dutch Health Care Institute (ZIN), regulatory information, and the company’s advertising materials, media presence, and website. All participants were informed of the respective roles and positions of each of the researchers involved in the process and the purpose of the research via the informed consent form, and were invited to ask questions both before and after each study inclusion. Social scientists with extensive ethnographic experience (first author, last author) led study design, data collection, and analysis; medical researchers with some ethnographic training contributed to data collection. Documents analyzed are available to the public. We used the SRQR checklist when writing our report.<sup>61</sup> see Appendix 1 for more details.

This study includes ethnographic observations and fieldnotes from several researchers in the context of a wider collaboration between social scientists at Erasmus School of Health Policy and Management (ESHPM) and medical researchers at EMC, producing a diversity of perspectives unavailable if all observations had been conducted by a single ethnographer. Our data were collected in a traditionally ethnographic manner: we pursued data collection iteratively and abductively in combination with data analysis, adding more data sources to fill gaps observed in previously-collected data sets.<sup>62</sup> This process, and the observational data itself, was given additional structure by our focus on legitimacy breakdowns. Nevertheless, a key component of ethnographic research is the capacity to allow new themes to emerge from spontaneous and unexpected data points; our focus on legitimacy breakdowns framed, rather than prescribed, the data collected.<sup>63</sup>

## Data collection

Data were collected over a period of three years from a wide variety of sources and then integrated as a means of studying various components of the larger system over time. Ethics approval was granted by Erasmus University Rotterdam (EUR) for ethnography and interviews (ETH2324-0010). Informed consent was collected and data was pseudonymized and stored securely to ensure adherence to European privacy guidelines.

*Collaborative ethnographic methods.* Some data were acquired within the context of a pilot study examining the

conditions and feasibility for a study of the introduction of the app into primary care in the Netherlands, conducted by EMC (MEC2021-0254).<sup>64</sup> This pilot feasibility study was part of an ongoing collaboration between EMC and ESHPM. Additional data were collected through focus groups and interviews with medical professionals such as doctors, nurses, and administrative staff. Participant observation took place in-person at GP practices.

Some data were also acquired in the context of a study investigating the diagnostic accuracy of the app in a clinical setting, conducted by EMC (MEC-2019-041).<sup>65</sup> The app was used to assess lesions after clinical diagnosis and before obtaining histopathology; the use of the app had no impact on patient care or health. Participants observed with malignant lesions were over age 50; student benign control participants were between ages 22-30. Participant observations took place both in-person at Dutch hospital clinics, and remotely via Teams or video recordings when in-person observation was not possible due to COVID-19 pandemic restrictions.

Convenience sampling was used in both studies: ethnographic data was collected from participants who were recruited based on clinical study criteria.<sup>64,65</sup> Accuracy Study participants were asked to participate in-person by dermatologists at their appointments for a suspicious skin lesion check; GPP participants were asked to participate by phone or email after contacting their GPs regarding a suspicious skin lesion. The total inclusion time for all participants was between 20 min and 45 min.

**App experiences ethnography study.** Focused on micro-level relationships (between users and the app), the App Experiences study (AE) provided data about the physical and practical management of the smartphone and the app by individual users, and how the app impacts the patient experience of assessment of suspicious lesions. App users were recruited through personal networks and the EUR via text message or email using the snowball sampling method. Participation lasted 30 min to one hour, and included a short interview and observation while the participant downloaded and registered with the app and used the app to make a risk assessment of a skin lesion of their choice. Participation occurred in private homes or in closed rooms at the EUR. Participants ( $n = 15$ ) ranged in age from 25 to 40 years and were roughly 50% Dutch, with the majority of foreigners originating from elsewhere in Europe. All but one participant had obtained a Masters degree or higher. Participants gave informed consent for participation and audio recording in writing and had the opportunity to comment on their answers. Researchers never recorded medical information provided to the app, nor did this study use or access app photos. This study was funded independently through the lead author's internal departmental funding and received ethical approval from ESHPM (ETH2324-0010).

Additionally, the lead author participated in the study through autoethnographic observations of three skin lesion checks using the app, a GP visit, and a dermatology follow-up. This data was recorded in detailed fieldnotes.

**Interviews with expert stakeholders.** The lead author conducted interviews with expert stakeholders working in research, health insurance, policymaking, and technology development ( $n = 8$ ). These semi-structured ethnographic interviews<sup>66</sup> lasted between 30 min and 2 h and were audio-recorded and transcribed. Some follow-up emails and documents provided during these interviews were also analyzed as data with participant permission (Table 1).

## Analysis

We abductively moved back and forth among various data sources and theoretical concepts<sup>67,68</sup> to account for the breadth and diversity of the data. We continued data collection until saturation was reached. All interviews and focus groups were recorded and transcribed. Observations were either recorded and transcribed or described in detailed fieldnotes;<sup>69</sup> if both existed for an interaction, these were collated before analysis. The final, fully digital data set (including transcriptions, fieldnotes, documents for analysis, recordings, and spreadsheets of metadata) was filed on secure EUR or EMC servers, according to GDPR standards.

The first author then abductively coded this data using thematic analysis techniques<sup>70</sup> in a combination of Microsoft Word, Atlas.ti, and Excel. The first and last authors then discussed the first round of coding and initial ideas generated from these codes.<sup>71</sup> For example, in one round, the first author coded practices of users that did not fully match the instructions on the company website; within these practices, the first author coded the emotional quality of the interaction. The first and last authors then discussed how these emotional qualities may relate to legitimacy. The first author then returned to the data to verify the fit of this idea before continuing to code. Using an iterative process with several more rounds of coding, idea generation, and regular feedback and discussion meetings with the last author, the first author continued thematic analysis until all legitimacy issues observed in the data had been adequately accounted for and described.

Given our approach to legitimacy as an inherently relational concept, we analyzed our data in the context of the particular relationships impacting the embedding of the app in the healthcare system. Within this relational context, breakdowns were especially visible in misalignments between narratives or expectations and practices. Finally, we explored the affective nature of these breakdowns, which was particularly observable in relationships between people and the app, and doctors and patients. A

**Table 1.** Data sources and researchers.

Type of inclusion	Conducted by	Number of participants	Date
Interviews with medical researchers	Author 1 (female)	2	2020
Interviews with app company staff	Author 1	3	2021
Autoethnography	Author 1	3 instances (study participant + independent high-risk rating)	2021, 2022
Interviews with policymakers and insurers	Author 1	3	2023
App Experiences (AE) ethnography study with individual users	Author 1	15	2023
GP Pilot Study (GPP) inclusions with ethnographic data	Authors 1, 2 (female), masters students (female, male)	31 + unspecified doctors + 2 researchers.	2022–2023
GPP focus groups	Authors 2, 6 (male)	GPs ( $n = 5$ ), doctor's assistants ( $n = 3$ )	2022
Ethnographic observations of GPP inclusions	Authors 1, 2, masters students	4 (2 patients, 1 researcher, 1 doctor)	2022
Accuracy study remote (teams) and in-person ethnography inclusions	Author 1	9 participants (3 remote), 1 researcher, 1 doctor	2021
Accuracy study video inclusions	Author 1, masters students (female, male)	9 (4 participants, 4 helpers, 1 researcher)	2021
Document analysis	Author 1	5 documents	2020–2023

visualization of the themes and subthemes produced through this analytical process can be found in Figure 1.

## Results

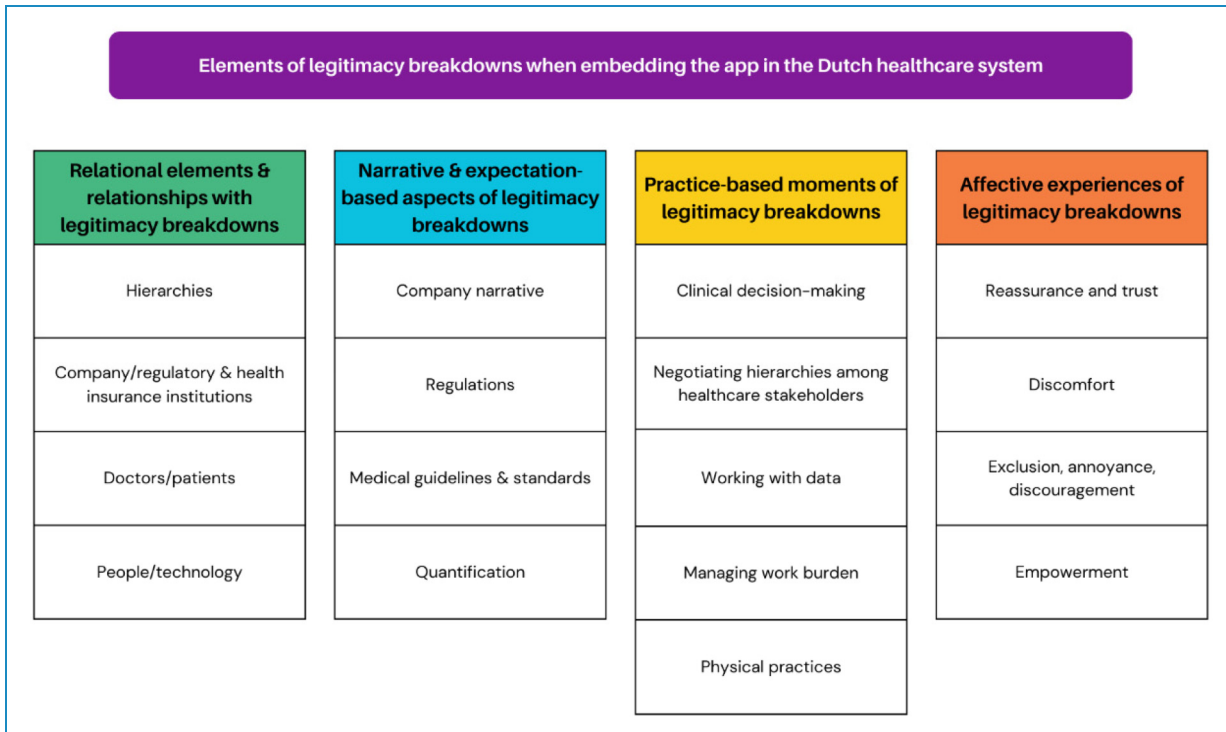
As we consider the embedding of new health technologies in practice through the analytical lens of legitimacy, we focus on the processes and activities through which such technologies gradually do or do not become taken-for-granted. Legitimacy can be understood as a process<sup>72,73</sup> that is shaped in the interactions between various stakeholders, the technology, the organizational context and the broader healthcare system requirements in which the technologies are developed. This also includes the narratives about the technology, systems of interaction already present in the context into which a technology is being introduced, and affective responses to the technology in use. We analyze “moments of breakdown” of legitimacy to understand what assumptions impact the process of embedding this new technology in the Dutch healthcare system.

Here, we consider moments of legitimacy breakdown for the app as it moves through different levels of the healthcare system in the Netherlands. Within our results, we gradually zoom in on relationships and interactions involving the app

from the macro level to the micro level<sup>74</sup> to guide our readers through the various themes and subthemes identified in our analysis. We start at a system-wide, macro level with a discussion of the company narrative in the context of the Dutch healthcare system. We then move toward a discussion of regulation and insurance practices, and how these practices negotiate between the company narrative and the Dutch health care context. From there, we move to a meso-level examination of medical practices in clinics, specifically between GPs and patients, and how the interaction of these practices with narratives and expectations of the app generates legitimacy breakdowns. Finally, we examine micro-level practices among individual users. We begin with a discussion of physical user practices, followed by an exploration of how expectations about the app, whether defined by the company narrative or other elements in the user's context, interact with user practices to generate an affective experience for individuals (Figure 2).

### *The company narrative in context: accuracy, risk and institutional positioning*

In the Netherlands, the current standard of care for skin cancer risk assessment runs through GP offices. Everyone



**Figure 1.** Themes and subthemes.



**Figure 2.** Results structure.

who lives or works in the Netherlands is legally required to take out a standard health insurance package to cover their medical costs.<sup>75</sup> 95% of Dutch residents are registered with a GP.<sup>75</sup> For a skin cancer risk assessment, a person with a suspicious skin lesion should schedule an appointment with their GP. Depending on the type of cancer suspected, the GP has several options: they can treat the lesion themselves, refer the patient to a dermatologist, or take a “wait and see” approach.<sup>76</sup>

This skin cancer risk assessment app is currently marketed to the entire Dutch population. In this model, individuals can download the app onto their own smartphones and register using their insurance information. They can then use the app without a doctor to perform a skin cancer risk assessment at home. If the app produces a “high risk” result, text on the app recommends that the user make an appointment with their doctor.

The company’s narrative describes the app as an efficient, effective, and institutionally-approved solution to an urgent healthcare problem. The company website presents “skin cancer” as a life-threatening risk that must be “caught in time” and the app as an important innovation for early detection that can provide a reliable assessment in “30 seconds” and is “95% accurate.” Institutional legitimacy is invoked through language and visuals: the term “regulated” appears four times on the “Explore SkinVision” page alone, “clinically validated” appears twice, and the website is plastered with icons denoting regulatory approvals and partnerships with large, trusted organizations.<sup>53</sup> By addressing legitimacy in terms of both medical scientific validity standards and institutional norms, the website appeals to doctors and policymakers. Simultaneously, the company addresses other audiences with an instrumentalist discourse common to start-up culture that positions AI mHealth as a positive, highly reliable change to an outdated healthcare system.<sup>77</sup> Combined with a classification of skin cancer as a healthcare

“emergency,” this discourse appeals to both potential users and investors interested in technological solutions to societal problems. For example:

The SkinVision app is empowered by a highly accurate AI algorithm and is supported by an advanced quality system involving the best **dermatologists** worldwide, which has only one goal: making sure you visit the **right doctor** at the first sign of risk for skin cancer” (emphasis added) (SkinVision website, accessed 6 May 2024)

Through an earlier mention of dermatologists, the company implies that dermatologists are “the right doctor” for initial skin cancer screening. Phrasing the problem in terms of “first sign of risk” rather than “first sign of cancer” outlines the company’s goal to become universal preventative care. This narrative suggests a threat-aware orientation to both skin cancer and the health-care system that requires individuals to be proactive to protect their own health, a common narrative around the introduction of new health technologies.<sup>78</sup>

In practice, however, skin cancer and the app’s performance appear significantly more nuanced than this narrative suggests. For instance, the vast majority of skin cancers are forms of basal cell carcinoma (BCC), which grows slowly and only rarely metastasizes.<sup>18</sup> These cancers are therefore not life-threatening, or even particularly “risky” from a health perspective, though they require treatment and their prevalence creates a significant burden on the care system.<sup>18</sup>

Additionally, the algorithm does not work well for people with darker skin tones<sup>79</sup> for a variety of reasons, both social and technical.<sup>80–82</sup> Company dermatologists review any low-contrast lesions photographed, which includes most lesions on dark skin. The app only assesses photos of lesions and takes symptoms of an individual lesion into account in its recommendations. Skin type and overall skin cancer risk are assessed through separate voluntary questionnaires for the user’s knowledge only; these do not inform the assessment of photographed lesions.<sup>83</sup> However, this is not necessarily clear within the app: two AE participants believed their lesion risk assessment was influenced by their skin type questionnaire (Participants 4, 7) before the app was changed to direct users to photograph their lesion before filling in the questionnaires in mid-November, 2023.

In terms of scientific validity, “95% accuracy” reflects the app’s 95% sensitivity rate in a clinical validation study.<sup>84</sup> However, according to this same study, the app’s specificity, an important part of “accuracy” in scientific terminology, is significantly lower at 78%.<sup>84</sup> This means that the app is relatively good at avoiding false negatives (i.e., rating a cancerous lesion as low-risk) but not as good at avoiding false positives (i.e., rating a non-cancerous lesion as high-risk). However, the reliability of these

numbers is somewhat contested<sup>85,86</sup> and the numbers do not reflect possible changes in positive and negative predictive values among different populations and settings.<sup>65</sup> Across several studies, the app has been shown to rate a significant number of benign lesions<sup>87,88</sup> as “high-risk”; this may lead to unnecessary excisions and care. In one study, researchers observed a six-fold increase in insurance claims for benign skin lesions with “high-risk” ratings from the app, and a two-fold increase in claims for benign lesions with “low-risk” ratings, compared to a control group receiving standard care.<sup>83</sup> It is therefore possible that using the app makes patients more likely to seek care regardless of the risk rating.

Despite a company narrative of broad institutional support, the institutional positioning of the app in practice is more ambiguous. On the one hand, the app has CE certification in compliance with the current EU-level regulations of medical devices (though this regulation will soon change<sup>89</sup>). Several large insurance companies in the Netherlands already reimburse the app in the basic package and other insurers include the app in supplementary insurance packages. After years of opposition to AI for skin cancer risk assessment, the Dutch Dermatology Association released a position statement in 2023 detailing possible benefits and drawbacks of various scenarios for AI mHealth apps but not approving all scenarios or the app’s current position in the care system.<sup>90</sup> The first randomized control trial that would provide evidence at EBM standards is currently in progress,<sup>91</sup> and the app does not yet have approval from ZIN (which is not unusual for a small company). While the app is clinically validated,<sup>65</sup> appropriate positioning of the app within the healthcare system is still unclear for both regulators and medical professionals.

### *Regulation and insurance practices*

While some regulations and insurance practices align well with the company narrative, significant deviations underscore the complexity of embedding a new technology into the healthcare system. Therefore, our presentation of data highlights the moments of legitimacy breakdown, rather than focusing on alignments.

In the Netherlands, health insurance is private but highly regulated. ZIN “supervises and stimulates the quality, accessibility, and affordability of healthcare in the Netherlands”.<sup>92</sup> Among other legal tasks, ZIN “advises the minister on whether care should be included in the basic benefit package of publicly funded health insurance; distributes public funds among health insurers based on risk equalization; and improves exchange of digital information between healthcare providers”.<sup>92</sup> While ZIN offers explicit recommendations for coverage under the basic insurance package for approximately 10% of new health-care services and technologies (including medicine), most of the time, healthcare actors, like insurers, decide on

their own what should be covered based on the Healthcare Act (Zorgverzekeringswet).<sup>93</sup>

Currently, preventative care is not covered by insurance with exceptions for some screenings and vaccinations based on risk status and/or age group. With few exceptions, GPs act as gatekeepers to all other healthcare, referring patients to specialists as needed. However, ZIN does not have the capacity to review every new health technology on the market. Therefore, discretion is left to insurers about what to include in the basic package when dealing with most new technologies, especially from smaller companies.

This is what has happened with this app. The app's current positioning as preventative care available to anyone on demand makes it an unusual case, and potentially problematic from a Dutch regulatory perspective. Despite lacking the institutional legitimacy of a "formal opinion" from ZIN, the app is reimbursed by several insurers for all members without a referral.

In the case of SkinVision, one of the larger healthcare insurers decided that they believed it was part of the basic healthcare insurance and they started paying from the basic healthcare. That means that they get a part of your income as an insurer from the insurance. But the major part comes from the government. They collect income tax and pay the healthcare providers for the care they delivered, and insurers get that money back based on the amount of healthcare insured (clients) received. But if you start sending invoices (for the app), you will get a larger share of the pie, so to say, and that's not good. It should be the same for everyone. (Health insurer 1)

Heavy regulation of insurance policies and tax-based contributions to healthcare costs make the app's current reimbursement status somewhat risky on both a legal and financial level if ZIN issues a formal opinion against inclusion in the basic package. Conversely, using other kinds of relationships, i.e., with insurance companies, to maintain legitimacy without official approval may be a successful strategy for a small company that needs income to continue to exist. Without the resources of giant pharmaceutical companies, these kinds of trade-offs can be the only affordable way for smaller tech companies to produce the kind of evidence necessary for formal approval by ZIN.<sup>94</sup> As it wrestles with the longer and more official avenues necessary to form legitimizing relationships with regulatory bodies in the Netherlands, the company (and the app by proxy) has developed some institutional legitimacy through relationships with a critical mass of insurers.

### GP practices

Moving toward the meso level of relationships, we now turn our attention to the ways in which the app disrupts standard care for skin cancer risk assessment in practice

at a clinical level. We focus here on GPs as the gatekeepers for all secondary care in the Dutch healthcare system,<sup>84</sup> and therefore, as the first point of contact for people who use the app in the Netherlands.

*Following the standard of care.* We first explore the "backstage"<sup>95</sup> of interactions between patients and GPs without technology involved. There is significantly more happening during a skin cancer assessment in a GP office than is captured through the app. In this section, we discuss what is under the surface of these interactions that is not captured by the technical narrative of the app's performance.

In the Dutch GP guideline for suspicious skin lesions,<sup>76</sup> the process for how to evaluate skin lesions and treat skin cancer at the GP level is clearly delineated. GPs are the first point of triaging. GPs should do a full body scan for patients presenting with a suspicious lesion. If the GP believes a lesion is precancerous or a low-risk basal cell carcinoma, they should remove it or treat it with methods like a cream or liquid nitrogen, depending on the lesion and location; lesions such as high-risk basal cell carcinoma,<sup>76</sup> squamous cell carcinoma, and melanoma should be referred to a dermatologist for treatment and removal. All excised lesions should be histopathologically evaluated to confirm the diagnosis.<sup>76</sup>

Guidelines can appear to be somewhat "algorithmic" on paper, in that they use a step-by-step process to accomplish a task,<sup>96,97</sup> and thus potentially translatable into computer-based reasoning systems. In practice, a GP's course of action for a suspicious lesion is determined through a conversation with the patient instead of an isolated judgement of the GP about the appearance of a particular lesion; the latter would be more analogous to an app-based risk assessment. These conversations between patients and GPs (GPP) usually encompass far more factors than the skin lesion itself:

The doctor listens carefully to the patient's questions and concerns. The patient tells her story and mentions again that she has had radiation therapy (for cancer). After this, the patient says that she's worried about a number of lesions on her skin, and motions to several lesions on her face and throat. She says her partner remarked on a lesion on her back as well. This represents more lesions than the patient had presented to the medical researcher using the app.

The doctor examines the lesions using the dermatoscope. The doctor says not to worry about the lesion on her throat, but she has doubts about the lesions on her face and back. Therefore, she wants to refer the patient to a dermatologist for further advice. The patient answers that she agrees with the doctor, and seems satisfied with the outcome. (GPP, R1 observation, practice 3)



In this case, the risk assessment from the app was “low.” However, in the diagnostic interaction, the doctor accounted for the location and number of lesions, and the patient’s medical history, in addition to lesion appearance and complaints. Interestingly, the patient had not presented this entire history, nor all of the lesions to the medical researcher and the app, but instead waited until she could speak to a doctor. In doing so, the patient framed her complaints in the terms most comprehensible to the actor (app or doctor) from whom she sought care.

*Privileging quantified data.* Most GPs stated that the app’s rating would not change their diagnosis. Some mentioned they would not change their diagnosis if they were “sure,” but they may turn to the app for “reassurance” when doubtful. However, in practice, the app may have a large impact on the GP’s treatment plan, regardless of how “sure” the GP felt about their diagnosis at the start:

The doctor examines the patient with a dermatoscope and thinks aloud. She says she suspects that it is an old-age wart. The patient says that a scab has fallen off on the skin spot. She used the app when the scab was still on and got a ‘high risk’ result. After the scab had fallen off, she used the app again. This time, she says that the result is ‘low risk’. I remember that the ‘high-risk’ result was the one recorded in the study’s system...Initially, the doctor says not to worry, but says if the app gives a high-risk result, she would refer the patient to a dermatologist. The doctor states that she assumes a ‘low risk’ result and indicates that she is still thinking the lesion is an old-age wart. However, she says, because of the earlier high-risk result from the app, she will schedule the patient for a follow-up appointment to remove the skin spot and send it for analysis. I am amazed that the doctor allows her plan to depend very much on the app result. The patient leaves the doctor’s office. I let the doctor fill in questions for the medical study on the iPad. Later, when I check the questionnaire again, I see the doctor has completed all questions. However, when asked whether she would adjust her plan based on the risk result, the doctor answered ‘No’. This contradicts the doctor’s earlier statements! (GPP, R2 observation, practice 2)

The influence of AI decision support systems on expert opinions is well documented.<sup>98,99</sup> In this case, it appears that AI influenced the GP more strongly in practice than she indicated in a survey. While this could be an example of survey user error<sup>100</sup> or concern for legal liability,<sup>101</sup> it is likely that the GP was either unaware of the influence or did not perceive it to be “influence,” resulting in the difference between observations and survey data. The app fits easily within pre-existing frameworks that are considered part of good evidence-based medical practice,<sup>102,103</sup> and thus benefits from privileges already accorded to

technology and quantification in the practice of medicine. These privileges allow the app to slot into pre-existing taken-for-granted ideas<sup>104</sup> about the superiority of quantified data over experiential data and qualitative expertise. This easy fit thus represents a moment of breakdown in legitimacy not for the app itself, but because the veracity of the app’s recommendations is so taken for granted that the app inadvertently subverts the existing supremacy of physicians’ expertise over technology recommendations. Rather than using the app as a tool for medical investigation in the way that the GP used the dermatoscope, in this interaction, the app’s recommendation took on increased importance. Whether this change resulted from unnoticed influence, or concerns about legal ramifications of contradicting a technological solution’s presumed-accurate recommendation, the app’s influence on the physician’s decision contradicts basic assumptions about medical hierarchies (that human judgment always takes precedence over AI-based decision support).

*Managing work burden.* GP practices also contradict assumptions made in both the company narrative and medical context about AI as work burden relief. Healthcare providers regularly suggested a primary interest in the app as a method of relieving the work burden from seeing too many patients. GPs and assistants (GPP) frequently cited anxious patients without good reason to seek care as a source of this extra burden. Some GPs saw this as an opportunity for the app:

The GP said that he would like to have the application during consultations for extra certainty in case of doubts about the diagnosis. He would also like it if patients could use the application at home beforehand. Since the app has a high sensitivity rating, if there is then a low risk, he would like to use a low-risk rating to prohibit patients from coming to the consultation hour. (GPP, Observation, practice 3)

Other GPs, however, thought the app would make no difference:

The hypochondriacs who have been here ten times with spots and things...like, yes, I can see that you’re very worried, but the app won’t help with that. I need to get them in the consultation room, provide some nuance. (GP, Focus group 1)

Some GPs noted that impatience when it comes to skin cancer is understandable, even when it puts pressure on the practice:

People don’t want to wait long for a spot. Many people have already googled, so don’t say you can only see the doctor

on Wednesday afternoon...I mean, that's quite a long time.  
(GP, focus group Practice 2)

However, there were no GPP participants who fit this “hypochondriac” description. While GPs were sometimes dismissive of a patient's concerns about a lesion, they never indicated that the patient should not have come in for a skin check. In interviews with researchers after a GPP inclusion, GPs always accounted for a reasonable source of an anxious patient's concern (for instance, a family history of melanoma). In several instances, the GP believed a lesion (or other lesions on the body) to be high risk and worth a referral to the dermatologist without patient request, even when the app provided a low-risk rating.

These observations suggest that expectations of work burden relief from the introduction of AI tools may not align with the experience of work burden in practice for GPs. While the relief of work burden is not explicitly promised by the app, the company narrative plays into basic assumptions about the use of technology in healthcare, bringing this issue to the forefront. This threat to legitimacy is therefore only made visible when the app is embedded within a workflow in practice, such as at a GP office, when expectations of work burden relief are not met in practice.

### User practices

In this section, we explore how practices of individual users dealing with the app reveal legitimacy breakdowns at the microlevel. These breakdowns represent misalignments with assumptions based on both the company narrative and the context of app use, in this case, the personal nature of individuals' smartphones.

People often struggle to use the app alone. Only a small minority of GPP participants were able to successfully use the app at home before coming in for their appointment. Even when assisted by a researcher, several participants were unable to use the app successfully at all.

The reasons for usage issues varied, but the location of the lesion was a primary factor. The app can only use the rear camera of a smartphone for quality reasons. As such, it was impossible for most people to center the lesion and hold the phone steady without the help of a partner, unless the lesion was on their forearms or another easily accessible body part. Lesions on the head universally required a partner's help.

Patients and helpers (including researchers trained to use the app) often physically struggled with the technology, trying with one hand before realizing they needed two to steady the phone in awkward positions. When the app still could not recognize the lesion, participants and helpers usually gave up after a few minutes of struggle; as the husband of a clinic user put it: “Well, we tried”

(Accuracy P4). In the Accuracy and GPP studies, medical researchers would step in at this point. In most cases, researchers were able to take the photo; in several GPP inclusions, however, the researcher also could not get the app to recognize the lesion.

While some participants in all studies were happy to ask for help, others made it clear that having to ask for help to photograph a lesion would be a major barrier to using the app:

One participant struggled for several minutes alone, despite knowing he could ask for help. “The hell I can't take this photo,” he said through gritted teeth. A few minutes later, he succeeded, after contorting himself uncomfortably to try to aim the camera correctly. He smiled and said proudly, “I did it, see, I'm a big boy.” Later, he clarified that he meant that if it's about his health, he wants to be fully responsible for it, including taking the photos; asking someone else to do something he feels he should be able to do alone was an absolute last resort. (AE P5)

In this case, the app transferred the responsibility<sup>105</sup> for cancer risk assessment away from the doctor and toward the user, whether because of the language and design of the app based on the company narrative, or because of contextual cues designating smartphones as extremely personal devices.<sup>106</sup> Even though this participant said that he has previously gotten lesions checked by a doctor, using the app felt like something he had to do alone. Several other AE participants (3, 9, 12, 13, 14, 15) either resisted help when offered or purposefully did not ask because they felt this was something they had to do alone.

*Affect in user-app relationships.* Finally, we zoom in on the affective elements in the interactions between individual users and the app. We analyze these issues as moments of breakdown driven by assumptions based on the company narrative and around expectations of how health technologies should work. Many of these breakdowns are produced by the inability of users to directly interrogate and negotiate with the app. In these moments, affective qualities such as “reassurance” or “annoyance” become more salient for users, but the same kind of interaction can have a very different affective quality depending on the app user and their past experiences with technology and medicine. Affective differences can therefore be insightful in understanding how the app becomes (de)legitimized in the context of specific kinds of interactions between the app, users, and medical professionals.

*Reassurance and trust.* All participants in the AE study assessed lesions that were either un concerning to the participant, or that had already been assessed by a doctor. All stated they would still go to a doctor directly if they were worried about a skin lesion; however, a few

participants did see the app as a potential replacement for doctors' visits in the future if they could trust the technology enough: "(It is) much better than having to go to the doctor. But I don't know, if at this point, this is reliable enough to avoid going to the doctor" (AE P13). Several mentioned that though the app may offer reassurance,<sup>107</sup> it would not convince them a concerning lesion was low-risk. The app is not considered a replacement for the doctor, and most participants echoed this sentiment. While they saw the app as a useful tool, they did not feel it was as trustworthy as a doctor's opinion.

If I go to the doctor with this rash, he sees your arm and he has this knowledge, and then he says, okay, yeah, well, let's go for another checkup or let's go to this. With the application, you have the feeling of agency about how to diagnose yourself. Yeah. If you do incorrectly, then you have the feeling that maybe the answer won't suffice, like, for example, with the skin color. And that changes if you would trust it, or if it would replace going to the doctor. (AE P7)

For this participant, the app is less trustworthy because it is mediated through his own lack of experience and expertise in both AI and skin cancer. He wanted an expert human involved. AE participants who received a teledermatology evaluation (2, 5, 14) found this reassuring; however even this human evaluation still did not replace the value of visiting a doctor if they were truly concerned about the lesion. The diagnostic interactions between doctors and patients are fundamentally different from risk assessment interactions between patients and an app; patients seek care in a medical interaction in ways that they do not in an interaction with a smartphone app. This implies that the app may not be perceived as a legitimate replacement for doctors' visits by patients because the service it provides lacks the human interaction and negotiation integral to doctors' visits, regardless of technical capabilities.

In the Accuracy and GPP studies, patients often explicitly stated that they did not trust the app and would always go to the doctor, regardless of the app's risk assessment. However, upon receiving a low-risk rating for a lesion, these same patients expressed visible relief (more relaxed physical demeanor, more talkative, smiling), with one explicitly saying, "That makes me feel better!" This phenomenon was noted outside of medical settings as well. An AE participant expressed strong doubts about the usefulness of the app prior to use:

The participant holds her phone too far away from her arm at first. The app does not recognize the lesion. "This is complicated," she says, "The flickering (check marks that appear when the camera is well-positioned) is panic-inducing, my banking app does this better." A few seconds later, the app recognizes the lesion and took the

photo. The app shows a blue screen while the photo was analyzed. "Huh, low-risk," she said, "Let's see what happens if I tell it it's changing." She aims the phone correctly on the first try, quickly takes another photo, and marks that her lesion has changed in the screen listing possible complaints. The blue screen reappears. "It's taking longer this time," she says, "It's like it's thinking more. That's reassuring somehow." She clarifies later that the reassurance comes from the impression that the app seems to be "listening" when she adds information. "Still low risk," she says, "Interesting. Ok, let's try another lesion, one that I've actually had concerns about." (AE P2)

While the participant still said she would not use the app when she had easy access to her GP, she said the app surprised her. She had proven the app was "not complete nonsense" to herself through practical use and she thought it may be helpful in certain contexts. Affectively, the app changed from "panic-inducing" tech junk to a mildly anthropomorphic AI device that could "think" throughout the course of her inclusion. This change in affect toward a somewhat reassuring quality also altered the legitimacy of the app in practice for her in ways that would have been impossible without personal interaction with the technology. For many participants, legitimacy was built through practical experience with the app.

**Discomfort.** The app includes two questionnaires about skin type and skin cancer risk which require qualitative judgments from users. Most users expressed significant discomfort with these questionnaires and worried that their answers may incorrectly impact the AI rating of individual lesions. These users initiated a conversation with the researcher (who was not a dermatologist nor connected to the company) about how to answer specific questions:

P4: I think that's quite difficult (to choose my skin color) without seeing an example. Like I know I'm not light brown or dark brown or black, but I guess...not ivory white? Maybe something like fair to pale, fair to beige?

Researcher: It's your call. It's all very subjective, right, people are being asked to rate their own skin.

P4: How does your skin respond to sun...That's also a bit difficult...burns moderately?

(AE P4)

Some tinkered<sup>108</sup> with the technical processes in the app itself, moving back and forth between qualitative questions about lesion symptoms and the final risk assessment while changing their answers to determine how various answers impacted the outcome (Participants 2, 3, 4, 5, and 8). The app's lack of capacity for the conversation resulted in an

affectively different experience that could not be equated to a visit with a doctor.

*Exclusion, annoyance, and discouragement.* People also struggled with the administrative set-up of the app. Several AE participants mentioned that if there hadn't been a researcher available to help, they would have just given up. Few GPP participants managed to download and set up the app with their personal details, despite several attempts, before arriving at the GP practice. This period of struggle had an affective quality; participants with less confidence around technology seemed to feel intimidated and excluded as they wrestled with the digital forms, and often gave up quickly even when aided by the researcher (GPP and AE). For one AE participant, medical administrative difficulties were negatively loaded due to personal experience:

All the parts of the admin stuff...are things that more and more I have a strong aversion to, especially for healthcare, because I see how much my parents are struggling with all their healthcare apps. Not remembering the passcodes, the one for my mom, for my dad, then the email is not the same: they are so lost, and it's so much more complicated than it should be, and it's anxiety inducing for them at their age, that (when starting the app) I'm thinking, "Oh God, one more." (AE P5)

For another participant with a disability that requires constant contact with the health care system, the administrative and medical sides of the app had almost reversed affects: the medical side produced affects of exclusion and annoyance, while the administrative side felt routine.

Participant 8 was business-like, calm while she held the phone with both hands to go through the app's sign-up administration, with almost a bored expression on her face. I asked if she was bored. She laughed and said, "It's like filling out another survey...because I used to do it every day. I was part of this trial, and twice a day, I had to do this, and score, like how was your day? Did you do this? How active were you?" She also had trouble getting the app to recognize her chosen lesion. She tried moving the hair on her forearm. She became very still and furrowed her brow, holding the camera steady for at least 10 seconds before declaring it wouldn't work with some frustration. I suggested she move the phone closer to her arm and tilt it. The app took a photo. I mentioned that most people move the phone around when the app doesn't work. She responded, "A lot of the apps have been designed for people who are able-bodied, expecting people to be able not to shake their hands...then there's this loading screen that looks like it needs to be steady in order to work, so that just reinforced it." When asked if this made her feel stressed, she responded, "It's very cynical, but I thought,

'Well, this is fitting, because nothing works in healthcare.' And I wanted to be a good participant." (AE P8)

A relationship marred by affective discouragement clearly impacts who is able to use the app as intended, making it less likely that groups with lower technological literacy or physical difficulties will be able to use the app.<sup>64</sup> In this case, an affective quality of exclusion becomes realized in practice when participants who feel excluded give up.

*Empowerment.* Given the well-documented evidence that women's health concerns tend to be taken less seriously by doctors for both structural and social reasons,<sup>109,110</sup> it is unsurprising that some women viewed this app as a self-advocacy tool when dealing with the medical establishment. While men and people with a family history of melanoma occasionally noted during interviews that the app could be used for self-advocacy, they did not propose using it this way themselves with their own doctors.

Several female AE participants under age 30 talked about the app as a stopgap against what they perceived as insufficient care from their GPs. For instance, one woman who had moved to the Netherlands from a highly medicalized healthcare system was interested in the app as a screening tool to bridge the gap between the care she offered in the Netherlands and the care she is used to at home:

Most of the time (being dismissed at the GP) is fine, but I think I'm used to a different system, where if you have a small complaint, they check a lot. I sometimes think, hmm, I'd like to have more information. (AE P4)

These women stated that they found using the app to be a far more pleasant and "validating experience" (AE P4) than going to the GP to get a suspicious lesion checked. This was due to both the language on the app that takes their concerns seriously even when providing a low-risk rating and the ability to use the app to avoid an unnecessary and potentially stressful in-person interaction with their doctor. These women saw potential benefit in using the influence of AI on expert opinion<sup>98,99</sup> to bolster their position with their doctors in order to be taken seriously.<sup>111</sup> As one AE participant put it, "It's like a legitimacy vending machine, spitting out numbers" (AE P15).

As in the example of AI influence on doctor decision-making, younger women using the app as a form of self-advocacy does not represent a breakdown in legitimacy for the app itself. In fact, the practice of self-advocacy aligns well with the company narrative. Instead, this is a breakdown of legitimacy precipitated by misalignment with assumptions about the hierarchy of the medical system. Here again, the app subverts the supremacy of a doctor's judgment, rebalancing power in the relationship

between the doctor and patient in a standard medical interaction.

### Summary

Our analytical lens of legitimacy has allowed us to use “moments of breakdown” within our ethnographic data to understand how the company narrative and existing context-based assumptions within the Dutch healthcare system (mis)align with practices involving the app. This focus made it possible to study the unquestioned and unnoticed elements in both the narrative around the technology and the medical system as a whole in relation to the practices and affective experiences of using the app. We identified three particular moments of breakdown that are instructive for understanding such (mis)alignments.

First, lack of (or slow acquisition of) institutional legitimacy through official channels required the company to pursue informal workarounds,<sup>112</sup> specifically, developing relationships with individual health insurers, that allowed the company to survive without funding equivalent to a major conglomerate. Second, quantification privilege<sup>113,114</sup> both increased and invisibilized the influence of the app on healthcare interactions, thereby disrupting assumptions about medical hierarchies and standard measures of quality in patient care. Third, limitations of flexibility and

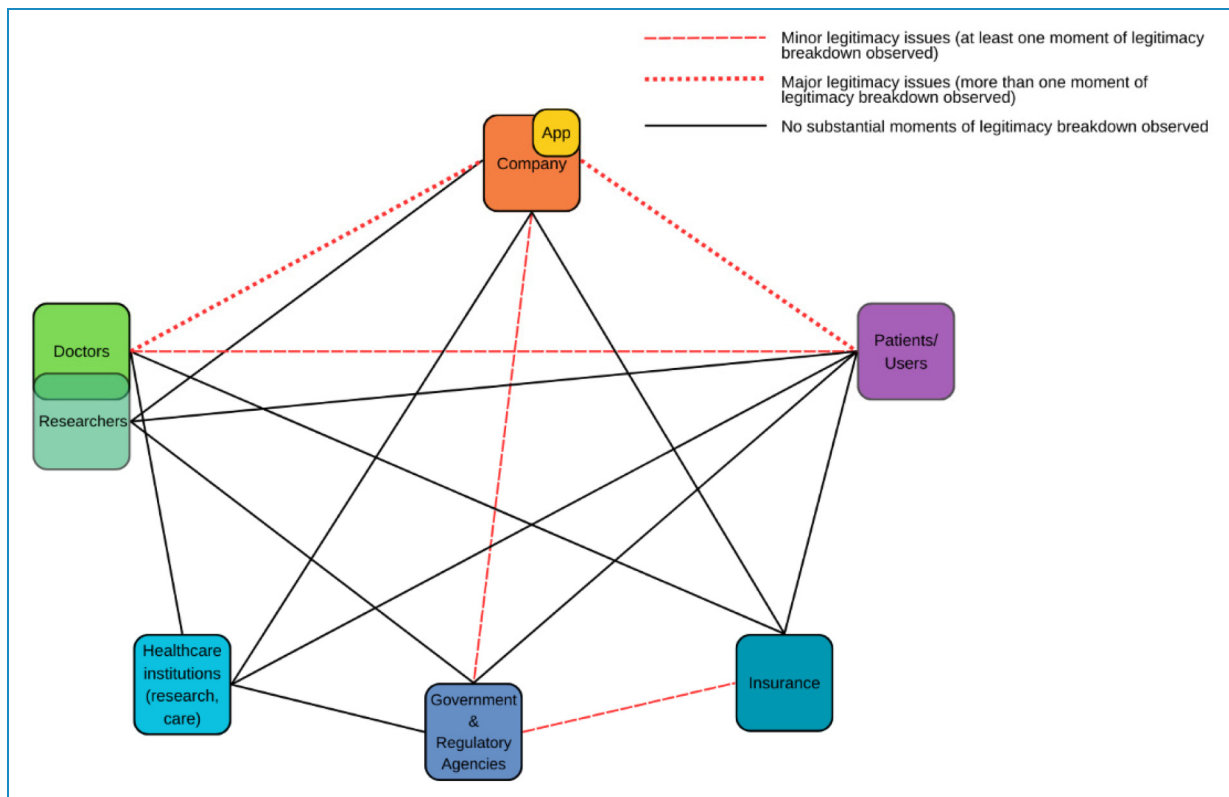
capacity for dialogue in interactions with the app (as compared with human doctors) led to unmet expectations around both work burden and ease of use because the app individualizes responsibility to the user, thereby sometimes creating additional work for both users and doctors. This limitation also increased feelings of exclusion for many vulnerable users.<sup>115</sup>

Below is a visualization of relationships impacting the embedding of the app in the healthcare system; based on our data, we show which relationships contained legitimacy breakdowns (Figure 3).

Relationships between the app/company and both doctors and patients/users contained the most moments of legitimacy breakdown. However, moments of legitimacy breakdown related to the app were also found between doctors and patients, the company and regulators, and regulators and insurers, and we did not find noticeable breakdowns in all legitimacy-impacting relationships.

### Discussion

In this ethnographic case study of an AI smartphone app for skin lesion cancer risk assessment, we used a legitimacy lens to understand how technology is (not) embedded within a healthcare system. We explored moments of legitimacy breakdown as a means of understanding the



**Figure 3.** Legitimacy-impacting relationships affecting the adoption, acceptance, and implementation of the app in the Dutch healthcare system.

assumptions that guide this embedding process. We conceptualized “legitimacy breakdowns” as moments when practices involving the app<sup>63</sup> and affective aspects of user/app interactions challenged the company narrative or other assumptions within the healthcare context. We then used this conceptualization to analyze the assumptions that guide the embedding of this app in the Dutch healthcare system at a macro, meso, and micro level. Here, we discuss the implications of these assumptions for embedding technology in healthcare, and our contributions to existing literature.

Our study contributes to the literature on legitimacy in health and technology<sup>103,116–119</sup> by revealing that practical interactions made the app’s utility and capabilities salient to certain users,<sup>22</sup> and in some cases, helped legitimate the app in situations where narratives were ineffective. This supports the findings of prior studies,<sup>34 35</sup> that suggest narratives alone are insufficient in understanding the full scope of how legitimacy is produced and maintained, instead conceptualizing legitimacy as a product of practices, discourse, and relationships. Our study contributes a focus on the physical and affective<sup>120</sup> negotiations that impact the legitimacy of a health technology to existing approaches that prioritize practices and relationships.

By examining differences between narratives and practices, policymakers and developers can target issues that will likely impact whether a technology is embedded in healthcare or not. Exploring the emotional qualities of user interactions with the technology and taking affective exclusion and discomfort seriously as indicators of user expectations can help developers find and understand issues early. Affective dissonance and non-user experiences are often ignored or marginalized in technology development, user testing, and policymaking around technology in healthcare,<sup>121,122</sup> however, as our data and others’ work shows, these dissonant opinions are invaluable to anticipation of challenges to technology embedding in healthcare.<sup>115,123</sup> On a more specific level, we also add to the existing literature on algorithm aversion in skin lesion risk assessment.<sup>124</sup> We provide a deeper explanation for why patients are consistently resistant to relying on AI rather than a doctor; namely that when narratives and assumptions about AI do not align with affective and practical experiences of AI users in medical practice, the legitimacy of AI as a replacement for human medical care is compromised.

Second, our study contributes to the literature on technology acceptance, adoption, implementation, and governance frameworks. These frameworks provide researchers with conceptual and methodological frameworks to explore specific embedding issues with attention to concepts like transparency,<sup>29,35</sup> trust,<sup>27,29</sup> and acceptability.<sup>22,23,28,31,33,125</sup> These frameworks are often used to inform decision-making not only for researchers, but for developers and policymakers. To avoid approaching acceptance, adoption, implementation, and governance challenges separately, some literature in this area also

covers more comprehensive approaches to technology embedding challenges, such as the NASSS framework<sup>9</sup> and Technology-Organization-Environment when used in healthcare.<sup>126</sup> Our finding that the process of embedding technology in healthcare is impacted by contradictions between narratives/assumptions and affects/practices builds on this work. This study supports the continued use of these more complex and nuanced frameworks and challenges the applicability of frameworks that focus solely on behavioral intention.<sup>25,26,31</sup> We add to these more comprehensive frameworks by presenting a legitimacy lens as a strategy for identifying possible embedding issues in advance that could allow researchers, developers, and other stakeholders to better anticipate challenges for embedding in practice from a socio-technical perspective.<sup>10,57,77,127</sup>

By foregrounding moments of breakdown in legitimacy through ethnographic data analysis, particularly of practice and affect,<sup>58,128</sup> studies of technology embedding in healthcare can access nuances that may be absent without this perspective. For instance, our results indicate that trust and legitimacy are not the same and that, while trust is an important component in the embedding of technology in healthcare, a healthcare technology may be able to obtain a level of legitimacy without it. This nuance has implications for researchers conducting studies about technology adoption, acceptance, implementation, and governance with a heavy focus on trust,<sup>27,29,35</sup> and suggests that legitimacy and trust may best be studied as separate criteria in studies of health technology embedding in general, despite their similarities.<sup>127,129</sup> This ethnographic approach to legitimacy could therefore also provide direction to pain points when exploring embedding problems.<sup>130</sup> This would encourage researchers to pursue studies centered on issues related to practical or affective dissonances with prevailing assumptions and help avoid study designs that take those assumptions for granted.

This study has several strengths: it represents the culmination of a long period (three years) of ethnographic data collection from many sources, including beyond a clinical setting. In clinical inclusions, some observations could have benefitted from more depth and richness; however, in an operational medical context, inclusions had to happen between other work. The composition of the team, similarly, represents a strength in that diverse perspectives allowed us to incorporate both medical and social science-informed understandings into our ethnography. This in turn required more coordination, potentially limiting the depth of some observations while broadening our overall perspective.

Because we were unable to conduct ethnographic research on-site with the company, we focused entirely on interactions between the company and outside stakeholders, rather than on the internal workings of the development, improvement, and implementation strategy at the company. While we believe these external interactions to

be most important for the production of legitimacy, we may have missed nuances in company strategy due to a lack of internal data.

A map of relationships impacting legitimacy (Figure 3) could be used by future researchers as a more structured way to explore embedding technologies in healthcare systems. Future scholars can apply and tailor this to their own studies of (digital health) technologies to search for legitimacy pain points within the relational contexts in which technologies operate. These maps may change over time as technologies move through the embedding process; while our data suggests that legitimacy conflicts may be limited to a relatively small number of relationships that are important to the embedding of a technology in healthcare (see Figure 3), this should not be taken to mean that all relationships without breakdowns observed are free from legitimacy issues. For instance, other studies have shown that within healthcare institutions, relationships among doctors and healthcare administrators can have a large impact on whether a change is implemented within a hospital,<sup>39</sup> though we did not observe dynamics like this in our data set, possibly because the app is marketed to individual users and insurance companies, rather than as a healthcare institution-based technology.

As a case study centered around a single technology being embedded in a single location, our study's specific findings may not be generalizable to other empirical studies of AI technologies in healthcare. However, this study has provided important insights into the utility of legitimacy as a lens to understand how technologies become embedded in healthcare and may be considered analytically generalizable,<sup>131</sup> in that our data supports and builds on findings from both the legitimacy literature and literature on embedding technology in healthcare. A legitimacy lens adds to comprehensive frameworks already in use and clearly identifies potential pain points.

## Conclusion

This study ethnographically investigated the embedding of a skin cancer risk assessment app in the Dutch healthcare system. We explored moments of legitimacy breakdown as a means of learning about the assumptions guiding the embedding of this AI-based application in the Dutch healthcare system.

We find that while many kinds of assumptions guide the embedding of the app, legitimacy breakdowns occur primarily when practices dissonate with an aligned combination of company narratives and context-based assumptions built into the healthcare system. These breakdowns occur at the macro, meso, and micro levels and have significant implications for the legitimacy of the app in the eyes of users, doctors, and the Dutch government, even when the breakdown occurs outside of direct relationships with the app (for instance, between doctors and

patients). These breakdowns have affective qualities for app users which, in turn, influence their (non-)use of the app.

Through this study, we have demonstrated that ethnographic exploration of legitimacy can be a powerful tool as we seek to better understand the acceptance, adoption, implementation, and governance of health technologies. We have shown that misalignments of practices and assumptions (narrative or context-based) within relationships at all levels of a healthcare system can challenge embedding goals for health technologies. New health technologies can disrupt existing hierarchies and require a renegotiation of the status quo; however, it is often unclear whether such renegotiations will lead to embedding of the technology or better health outcomes in the healthcare system. A legitimacy lens combined with an ethnographic approach addresses this issue by allowing researchers to pinpoint (anticipated) embedding problems and provide a clear yet nuanced perspective on the relationships impacting legitimacy of a health technology in context.

**Acknowledgements:** Many people made this project possible, but we could not have done it without our participants. For input and support during both data collection and writing, we would especially like to thank Ellen Algera, Renee Michels, Maura Leusder, Mixed Methods Anonymous, Tobias Sangers, Mayra Mohammad, Jeroen Revelman, Romy Litjens, and the Health Care Governance department at ESHPM.

**Contributorship:** Sydney Howe contributed to conceptualization (lead), data curation (lead), investigation (lead), methodology (lead), project administration (AE study and overall ethnography project), visualization, writing—original draft (lead), and writing—review and editing (lead). Anna Smak Gregoor contributed to data curation (supporting), investigation (supporting), and writing—review and editing (supporting). Carin Uyl-de Groot contributed to supervision (supporting), and funding acquisition. Marlies Wakkee contributed to project administration (GPP & accuracy studies), funding acquisition (supporting), resources (supporting), and writing—review and editing (supporting). Tamar Nijsten contributed to funding acquisition and resources. Rik Wehrens contributed to conceptualization (supporting), supervision (lead), methodology (supporting), investigation (supporting), and writing—review and editing (supporting).



**Declaration of conflicting interests:** The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The Erasmus MC Department of Dermatology has received an unrestricted research grant from SkinVision B.V. Dr Marlies Wakkee and Prof. Dr Tamar Nijsten serve on the Science and Dermatology advisory board for SkinVision. None of the authors received any direct fees for consulting or salary from SkinVision. SkinVision was not involved in the design of the study, data collection, data analysis, data interpretation, or writing of the manuscript. In addition, SkinVision was not involved in the decision to submit this work for publication.

The authors from Erasmus School of Health Policy and Management (Sydney Howe, Prof. Carin Uyl-de Groot, and Dr Rik Wehrens) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval:** Ethics approval was granted by Erasmus University Rotterdam for ethnography (App Experiences Study) and all interviews (ETH2324-0010). General Practitioner Pilot Feasibility Study (AIM HIGH): The study was assessed by the Medical Ethics Committee of the Erasmus University Medical Center (MEC2021-0254). They deemed it as not under the scope of the Medical Research Involving Human Subjects Act (WMO) and exempted it from further ethical approval. Written informed consent was obtained for all participants.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Erasmus School of Health Policy and Management and the Health Technology Assessment Department through the first author's PhD grant; and SkinVision B.V. (unrestricted funding for Erasmus MC Department of Dermatology).

**Consent to participate:** Written informed consent was collected from all study participants. Data was pseudonymized and stored securely according to EMC and ESHPM guidelines to ensure adherence to European privacy law.

**ORCID iDs:** Sydney Howe  <https://orcid.org/0000-0002-8656-8670>  
Anna Smak Gregoor  <https://orcid.org/0000-0002-7997-8532>

**Supplemental material:** Supplemental material for this article is available online.

## Note

a. Henceforth, we use the “the company” to refer to SkinVision as a company, and “the app” to refer to the SkinVision skin cancer risk assessment smartphone application.

## References

1. Angehrn Z, Haldna L, Zandvliet AS, et al. Artificial intelligence and machine learning applied at the point of care. *Front Pharmacol* 2020; 11: 759. 20200618.
2. Grote T and Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics* 2020; 46: 205–211. 20191120.
3. Smith H. Clinical AI: opacity, accountability, responsibility and liability. *AI Soc* 2020; 36: 535–545.
4. Zaidan AA, Zaidan BB, Albahri OS, et al. A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: coherent taxonomy, open issues and recommendation pathway solution. *Health Technol (Berl)* 2018; 8: 223–238.
5. Malhi IS and Yiu ZZN. Algorithm-based smartphone apps to assess risk of skin cancer in adults: critical appraisal of a systematic review. *Br J Dermatol* 2021; 184: 638–639. 20201214.
6. Ahuja G, Tahmazian S, Atoba S, et al. Exploring the utility of artificial intelligence during COVID-19 in dermatology practice. *Cutis* 2021; 108: 71–72.
7. Cadario R, Longoni C and Morewedge CK. Understanding, explaining, and utilizing medical artificial intelligence. *Nat Hum Behav* 2021; 5: 1636–1642. 20210628.
8. Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med* 2020; 260: 113172. 20200715.
9. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res* 2017; 19: e367. 20171101.
10. Cresswell K and Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. *Int J Med Inform* 2013; 82: e73–e86. 20121109.
11. Varabyova Y, Blankart CR, Greer AL, et al. The determinants of medical technology adoption in different decisional systems: a systematic literature review. *Health Policy* 2017; 121: 230–242. 20170124.
12. Roth LM. Defensive versus evidence-based medical technology: liability risk and electronic fetal monitoring in low-risk births. *Soc Sci Med* 2023; 317: 115565. 20221126.
13. Varela-Lema L, Ruano-Ravina A, Mota TC, et al. Post-introduction observation of healthcare technologies after coverage: the spanish proposal. *Int J Technol Assess Health Care* 2012; 28: 285–293.
14. Seyyed-Kalantari L, Zhang H, McDermott MB, et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med* 2021; 27: 2176–2182.
15. Adjekum A, Blasimme A and Vayena E. Elements of trust in digital health systems: scoping review. *J Med Internet Res* 2018; 20: e11254.
16. Du-Harpur X, Watt FM, Luscombe NM, et al. What is AI? Applications of artificial intelligence to dermatology. *Br J Dermatol* 2020; 183: 423–430. 20200329.
17. Hogarty DT, Su JC, Phan K, et al. Artificial intelligence in dermatology—where we are and the way to the future: a review. *Am J Clin Dermatol* 2020; 21: 41–47.
18. Schreuder K, Hollestein L, Nijsten TEC, et al. A nationwide study of the incidence and trends of first and multiple basal cell carcinomas in the Netherlands and prediction of future incidence. *Br J Dermatol* 2022; 186: 476–484. 20211227.
19. Sangers TE, Wakkee M, Kramer-Noels EC, et al. Views on mobile health apps for skin cancer screening in the general population: an in-depth qualitative exploration of perceived barriers and facilitators. *Br J Dermatol* 2021; 185: 961–969. 20210705.
20. Maier T, Kulichova D, Schotten K, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. *J Eur Acad Dermatol Venereol* 2015; 29: 663–667. 20140804.



21. Hengst TM, Lechner L, Dohmen D, et al. The facilitators and barriers of mHealth adoption and use among people with a low socio-economic position: a scoping review. *Digit Health* 2023; 9: 20552076231198702. 20230905.
22. Graeber J, Warmerdam E, Aufenberg S, et al. Technology acceptance of digital devices for home use: qualitative results of a mixed methods study. *Digit Health* 2023; 9: 20552076231181239. 20230613.
23. Virtanen L, Kaihlanen AM, Kainiemi E, et al. Patterns of acceptance and use of digital health services among the persistent frequent attenders of outpatient care: a qualitatively driven multimethod analysis. *Digit Health* 2023; 9: 20552076231178422. 20230525.
24. Schoville RR and Titler MG. Guiding healthcare technology implementation: a new integrated technology implementation model. *Comput Inform Nurs* 2015; 33: 99–107.
25. Nadal C, Sas C and Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. *J Med Internet Res* 2020; 22: e17256. 20200706.
26. Holden RJ and Karsh BT. The technology acceptance model: its past and its future in health care. *J Biomed Inform* 2010; 43: 159–172. 20090715.
27. Fox G and Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. *Inf Syst J* 2018; 28: 995–1019.
28. Dwivedi YK, Shareef MA, Simintiras AC, et al. A generalised adoption model for services: a cross-country comparison of mobile health (m-health). *Gov Inf Q* 2016; 33: 174–187.
29. Reddy S, Allan S, Coghlan S, et al. A governance model for the application of AI in health care. *J Am Med Inform Assoc* 2020; 27: 491–497.
30. May C and Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology* 2009; 43: 535–554.
31. Yarbrough AK and Smith TB. Technology acceptance among physicians: a new take on TAM. *Med Care Res Rev* 2007; 64: 650–672. 20070823.
32. Yuan CT, Nembhard IM and Kane GC. The influence of peer beliefs on nurses' use of new health information technology: a social network analysis. *Soc Sci Med* 2020; 255: 113002. 20200422.
33. Bramo SS, Desta A and Syedda M. Acceptance of information communication technology-based health information services: exploring the culture in primary-level health care of South Ethiopia, using Utaut Model, Ethnographic Study. *Digit Health* 2022; 8: 20552076221131144. 20221019.
34. Greenhalgh T, Hinder S, Stramer K, et al. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. *Br Med J* 2010; 341: 1091.
35. Licht KD and Licht JD. Artificial intelligence, transparency, and public decision-making why explanations are key when trying to produce perceived legitimacy. *AI & Soc* 2020; 35: 917–926.
36. Hornberger J. Who is the fake one now? Questions of quackery, worldliness and legitimacy. *Crit Pub Health* 2019; 29: 484–493.
37. Barker KK. Listening to Lyrica: contested illnesses and pharmaceutical determinism. *Soc Sci Med* 2011; 73: 833–842.
38. Reay T, Goodrick E, Casebeer A, et al. Legitimizing new practices in primary health care. *Health Care Manage Rev* 2013; 38: 9–19.
39. Sheard L, Marsh C, O'Hara J, et al. The patient feedback response framework - understanding why UK hospital staff find it difficult to make improvements based on patient feedback: a qualitative study. *Soc Sci Med* 2017; 178: 19–27.
40. Elshaug AG, Rosenthal MB, Lavis JN, et al. Levers for addressing medical underuse and overuse: achieving high-value health care. *Lancet* 2017; 390: 191–202.
41. Suchman MC. Managing legitimacy – strategic and institutional approaches. *Acad Manage Rev* 1995; 20: 571–610.
42. Hurd I. Legitimacy and authority in international politics. *Int Organ* 1999; 53: 379–408.
43. Haack P, Pfarrer MD and Scherer AG. Legitimacy-as-Feeling: how affect leads to vertical legitimacy spillovers in transnational governance. *J Manage Stud* 2014; 51: 634–666.
44. Weber M. *Economy and society: A new translation*. Cambridge, MA: Harvard University Press, 2019.
45. Kitchener M. Mobilizing the logic of managerialism in professional fields: the case of academic health centre mergers. *Organ Stud* 2002; 23: 391–420.
46. Sapir A. Mythologizing the story of a scientific invention: constructing the legitimacy of research commercialization. *Organ Stud* 2020; 41: 799–820.
47. Jain S and Ahlstrom D. Technology legitimacy and the legitimacy of technology: the case of chronic kidney disease therapies. *J Eng Tech Manage* 2021; 62.
48. Barnreuther S. Innovations 'Out of Place': controversies over IVF beginnings in India between 1978 and 2005. *Med Anthr Cross Cult Stud Health Illn* 2016; 35: 73–89.
49. Borup M, Brown N, Konrad K, et al. The sociology of expectations in science and technology. *Technol Anal Strateg Manag* 2006; 18: 285–298.
50. NN. OHRA, <https://www.nn-group.com/our-company/who-we-are/our-brands.htm#:~:text=OHRA%20is%20a%20direct%20insurer,employees%20serving%20about%20800%2C000%20customers> (2024, accessed 5 May 2024).
51. Kruis Z. 6 Reasons to Believe in Zilveren Kruis, <https://www.zilverenkruis.nl/en/consumers/insurances> (2024, accessed 5 May 2024).
52. CZ. Fieten en Cijfers, <https://www.cz.nl/over-cz/wie-wij-zijn/feiten-en-cijfers> (2024, accessed 5 May 2024).
53. B.V. S. SkinVision–Skin cancer melanoma detection app, <https://www.skinvision.com/> (2021, accessed 6 May 2024).
54. van Voorst R and Ahlin T. Key points for an ethnography of AI: an approach towards crucial data. *Hum Soc Sci Commun* 2024; 11: 1–5.
55. Star SL and Ruhleder K. Steps toward an ecology of infrastructure: design and access for large information spaces. *Inf Syst Res* 1996; 7: 111–134.
56. Haldar M, Engebretsen E and Album D. Legitimizing the illegitimate: how doctors manage their knowledge of the prestige of diseases. *Health* 2016; 20: 559–577.
57. Bagge-Petersen CM, Langstrup H, Larsen JE, et al. Critical user-configurations in mHealth design: how mHealth-app design practices come to bias design against chronically ill

- children and young people as mHealth users. *Digit Health* 2022; 8: 20552076221109531. 20220615.
58. Gjodsbol IM, Ringgaard AK, Holm PC, et al. The robot butler: how and why should we study predictive algorithms and artificial intelligence (AI) in healthcare? *Digit Health* 2024; 10: 20552076241241674. 20240324.
  59. Seaver N. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data Soc* 2017; 4.
  60. Weiss MC and Sutton J. The changing nature of prescribing: pharmacists as prescribers and challenges to medical dominance. *Sociol Health Illn* 2009; 31: 406–421.
  61. O'Brien BC, Harris IB, Beckman TJ, et al. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014; 89: 1245–1251.
  62. Geertz C. Deep hanging out. *New York Rev Books* 1998; 45: 69–72.
  63. Nicolini D. Zooming in and out: studying practices by switching theoretical lenses and trailing connections. *Organ Stud* 2010; 30: 1391–1418.
  64. Smak Gregoor AM, Sangers TE, Eekhof JA, et al. Artificial intelligence in mobile health for skin cancer diagnostics at home (AIM HIGH): a pilot feasibility study. *EClinicalMedicine* 2023; 60: 102019. 20230525.
  65. Sangers T, Reeder S, van der Vet S, et al. Validation of a market-approved artificial intelligence mobile health app for skin cancer screening: a prospective multicenter diagnostic accuracy study. *Dermatology* 2022; 238: 649–656.
  66. Spradley JP. *The Ethnographic Interview*. Long Grove, Ill: Waveland Press, 2016.
  67. Wehrens R, Stevens M, Kostenzer J, et al. Ethics as discursive work: the role of ethical framing in the promissory future of data-driven healthcare technologies. *Science, Technology, & Human Values* 2021; 48: 606–634.
  68. Tavory I and Timmermans S. *Abductive analysis: Theorizing qualitative research*. Chicago, Ill: University of Chicago Press, 2014.
  69. Spradley JP. *Participant observation*. Long Grove, Ill: Waveland Press, 2016.
  70. Clarke V and Braun V. Thematic analysis. *J Posit Psychol* 2017; 12: 297–298.
  71. Van Maanen J. Ethnography as work: some rules of engagement. *J Manage Stud* 2011; 48: 218–234.
  72. Suddaby R, Bitektine A and Haack P. Legitimacy. *Academy of Management Annals* 2017; 11: 451–478.
  73. Girschik V. Managing legitimacy in business-driven social change: the role of relational work. *J Manage Stud* 2020; 57: 775–804.
  74. Bitektine A and Haack P. The “macro” and the “micro” of legitimacy: toward a multilevel theory of the legitimacy process. *Acad Manage Rev* 2015; 40: 49–75.
  75. Wammes JS. Niek; Western, Gert International Healthcare System Profiles: the Netherlands, <https://www.commonwealthfund.org/international-health-policy-center/countries/netherlands> (2020, 2024).
  76. Baaten GGG BP, Damen Z, De Haas ERM, et al. NHG-Standaard: Verdachte huidafwijkingen (M108). *Nederlands Huisartsen Genootschap*. <https://richtlijnen.nhg.org/standaarden/verdachte-huidafwijkingen>. (2017, accessed 5 May 2024)
  77. Stevens M, Wehrens R and de Bont A. Conceptualizations of Big Data and their epistemological claims in healthcare: A discourse analysis. *Big Data Soc* 2018; 5.
  78. Wentzer HS and Bygholm A. Narratives of empowerment and compliance: Studies of communication in online patient support groups. *Int J Med Inf* 2013; 82: e386–e394.
  79. Astner S and Anderson RR. Skin prototypes 2003. *J Invest Dermatol* 2004; 122: xxx–xxxii.
  80. Gloster HM Jr. and Neal K. Skin cancer in skin of color. *J Am Acad Dermatol* 2006; 55: 741–760; quiz 761–744.
  81. Kundu RV. Practice gaps. Barriers in melanoma detection in ethnic minorities: comment on anatomic distribution of malignant melanoma on the non-Hispanic black patient, 1998–2007. *Arch Dermatol* 2012; 148: 801–802.
  82. Lester JC, Clark L Jr., Linos E, et al. Clinical photography in skin of colour: Tips and best practices. *Br J Dermatol* 2021; 184: 1177–1179. 20210318.
  83. Smak Gregoor AM, Sangers TE, Bakker LJ, et al. An artificial intelligence based app for skin cancer detection evaluated in a population based setting. *NPJ Digit Med* 2023; 6: 90. 20230520.
  84. Udrea A, Mitra GD, Costea D, et al. Accuracy of a smart-phone application for triage of skin lesions based on machine learning algorithms. *J Eur Acad Dermatol Venereol* 2020; 34: 648–655. 20191008.
  85. Freeman K, Dinnes J, Chuchu N, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *Br Med J* 2020; 368: m127. 20200210.
  86. Deeks JJ, Dinnes J and Williams HC. Sensitivity and specificity of SkinVision are likely to have been overestimated. *J Eur Acad Dermatol Venereol* 2020; 34: e582–e583. 20200622.
  87. Jahn AS, Navarini AA, Cerminara SE, et al. Over-detection of melanoma-suspect lesions by a CE-Certified smartphone app: Performance in comparison to dermatologists, 2D and 3D convolutional neural networks in a prospective data set of 1204 pigmented skin lesions involving patients' perception. *Cancers (Basel)* 2022; 14. 20220807.
  88. Freeman K, Dinnes J, Chuchu N, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ* 2020; 368: m127. doi: 10.1136/bmj.m127
  89. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *Br Med J* 2020; 368.
  90. Council EPA. REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. In: Council TEPA, (ed.). 2017/745. Brussels 2017.
  91. Wakkee MMM. Standpunt Artificial Intelligence in Mobile Health (mHealth). In: Oncologie D (ed.). Nederlandse Vereniging voor Dermatologie en Venereologie, 2023.
  92. Nijsten T and Wakkee M. Spot-studie (protocol web page), <https://www.erasmusmc.nl/nl-nl/kankerinstituut/patientenzorg/trials/spot-studie#5783694e-0ec3-41a7-8d67-8952d677b865> (2021).

93. van de Sande J, de Graaff B, Delnoij D, et al. Incorporating public values through multiple accountability: a case study on quality regulation of emergency care in The Netherlands by an independent regulatory agency. *Adm Soc* 2021; 54: 1178–1206.
94. Zorginstituut. Zvw-algemeen: Hoe werkt de Zorgverzekeringswet?, <https://www.zorginstituutnederland.nl/Verzekerde+zorg/zvw-algemeen-hoe-werkt-de-zorgverzekeringswet> (n.d., accessed 18 June 2024).
95. Zevestoski S, Brown P, McCormick S, et al. Patient activism and the struggle for diagnosis: gulf war illnesses and other medically unexplained physical symptoms in the US. *Soc Sci Med* 2004; 58: 161–175.
96. Carboni C, Wehrens R, van der Veen R, et al. Conceptualizing the digitalization of healthcare work: a metaphor-based critical interpretive synthesis. *Soc Sci Med* 2022; 292: 114572. 2021/11/29.
97. Merriam-Webster.com. Algorithm. *Merriam-Webster Dictionary*. 2024.
98. Ziewitz M. A not quite random walk: Experimenting with the ethnomethods of the algorithm. *Big Data Soc* 2017; 4.
99. Bansal G, Nushi B, Kamar E, et al. Beyond accuracy: the role of mental models in human-AI team performance. In: *Proceedings of the AAAI conference on human computation and crowdsourcing*, Stevenson, Washington, USA. 28 October–30 October 2019, pp.2–11.
100. Green B and Chen Y. Algorithm-in-the-loop decision making. *Proceedings of the AAAI Conference on Artificial Intelligence* 2020; 34: 13663–13664.
101. Gobo G and Mauceri S. *Constructing Survey Data: An Interactional Approach*. London: Sage Publications Ltd, 2013.
102. Martinho A, Kroesen M and Chorus C. A healthy debate: exploring the views of medical doctors on the ethics of artificial intelligence. *Artif Intell Med* 2021; 121: 102190. 20211012.
103. Stevens M, Wehrens R and de Bont A. Epistemic virtues and data-driven dreams: on sameness and difference in the epistemic cultures of data science and psychiatry. *Soc Sci Med* 2020; 258: 113116. 20200617.
104. Perrotta M and Geampna A. The trouble with IVF and randomised control trials: Professional legitimisation narratives on time-lapse imaging and evidence-informed care. *Soc Sci Med* 2020; 258.
105. McKnight B and Zietsma C. Finding the threshold: a configurational approach to optimal distinctiveness. *J Bus Venturing* 2018; 33: 493–512.
106. Schwennesen N. When self-tracking enters physical rehabilitation: from ‘pushed’ self-tracking to ongoing affective encounters in arrangements of care. *Digit Health* 2017; 3: 2055207617725231. 20170830.
107. Marchant C and O’Donohoe S. Homo prostheticus? Intercorporeality and the emerging adult-smartphone assemblage. *Inf Technol People* 2019; 32: 453–474.
108. Liu PL, Ye JF, Ao HS, et al. Effects of electronic personal health information technology on American women’s cancer screening behaviors mediated through cancer worry: differences and similarities between 2017 and 2020. *Digit Health* 2023; 9: 20552076231185271.
109. Mol A, Moser I and Pols J. *Care in practice: On tinkering in clinics, homes and farms*. Bielefeld, NRW: Transcript Verlag, 2010.
110. Perez CC. *Invisible women: Data bias in a world designed for men*. New York, NY: Abrams, 2019.
111. Henry C, Ekeroma A and Filoche S. Barriers to seeking consultation for abnormal uterine bleeding: systematic review of qualitative research. *BMC Womens Health* 2020; 20: 123. 20200612.
112. Chopra S, Zehrung R, Shanmugam TA, et al. Living with uncertainty and stigma: self-experimentation and support-seeking around polycystic ovary syndrome. In: *Proceedings of the 2021 CHI conference on human factors in computing systems*, Yokohama, Japan (virtual). 8 May–13 May 2021, pp.1–18.
113. Kannan-Narasimhan R. Organizational ingenuity in nascent innovations: gaining resources and legitimacy through unconventional actions. *Organ Stud* 2014; 35: 483–509.
114. Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. *Cult Health Sex* 2015; 17: 440–453. 20140611.
115. Ruckenstein M and Schüll ND. The datafication of health. *Annu Rev Anthropol* 2017; 46: 261–278.
116. Neven L. ‘But obviously not for me’: robots, laboratories and the defiant identity of elder test users. *Sociol Health Illn* 2010; 32: 335–347.
117. van Dijk S, Berends H, Jelinek M, et al. Micro-Institutional affordances and strategies of radical innovation. *Organ Stud* 2011; 32: 1485–1513.
118. Ruef M and Scott WR. A multidimensional model of organizational legitimacy: hospital survival in changing institutional environments. *Adm Sci Q* 1998; 43: 877–904.
119. Granja R and Machado H. Forensic DNA phenotyping and its politics of legitimisation and contestation: Views of forensic geneticists in Europe. *Soc Stud Sci* 2020; 53: 850–868.
120. Busfield J. The moral economy of the pharmaceutical industry: legitimising prices. *Health (London)* 2021; 25: 271–287. 20191017.
121. Navaro-Yashin Y. Affective spaces, melancholic objects: ruination and the production of anthropological knowledge. *Journal of the Royal Anthropological Institute* 2009; 15: 1–18.
122. Boon W and Edler J. Demand, challenges, and innovation. Making sense of new trends in innovation policy. *Science and Public Policy* 2018; 45: 435–447.
123. Wyatt S. Non-users also matter: the construction of users and non-users of the Internet. Oudshoorn N and Pinch T. *How users matter: the co-construction of users and technologies..* Cambridge, Mass: MIT Press, 2003, pp.67–79.
124. Jones M and Pietila I. Alignments and differentiations: people with illness experiences seeking legitimate positions as health service developers and producers. *Health* 2020; 24: 223–240.
125. Gaube S, Biebl I, Engelmann MKM, et al. Comparing preferences for skin cancer screening: ai-enabled app vs dermatologist. *Soc Sci Med* 2024; 349: 116871. 2024 0415.

126. Rouidi M, Elouadi A and Hamdoune A. Acceptance and use of telemedicine technology by health professionals: development of a conceptual model. *Digit Health* 2022; 8: 20552076221081693. 20220221.
  127. Yang J, Luo B, Zhao C, et al. Artificial intelligence health-care service resources adoption by medical institutions based on TOE framework. *Digit Health* 2022; 8: 20552076221126034. 20221005.
  128. Mathieu-Fritz A and Gaglio G. À la recherche des configurations sociotechniques de la télémédecine: revue de littérature des travaux de sciences sociales. *Réseaux* 2018; 207: 27–63.
  129. Lupton D. How does health feel? Towards research on the affective atmospheres of digital health. *Digit Health* 2017; 3: 2055207617701276. 20170410.
  130. Le G. Trading legitimacy: everyday corruption and its consequences for medical regulation in southern Vietnam. *Med Anthropol Q* 2013; 27: 453–470.
  131. Cleemput I, Devriese S, Kohn L, et al. What does the public want? Structural consideration of citizen preferences in health care coverage decisions. *MDM Policy Pract* 2018; 3: 2381468318799628. 20180925.
  132. Halkier B. Methodological practicalities in analytical generalization. *Qual Inq* 2011; 17: 787–797.
-