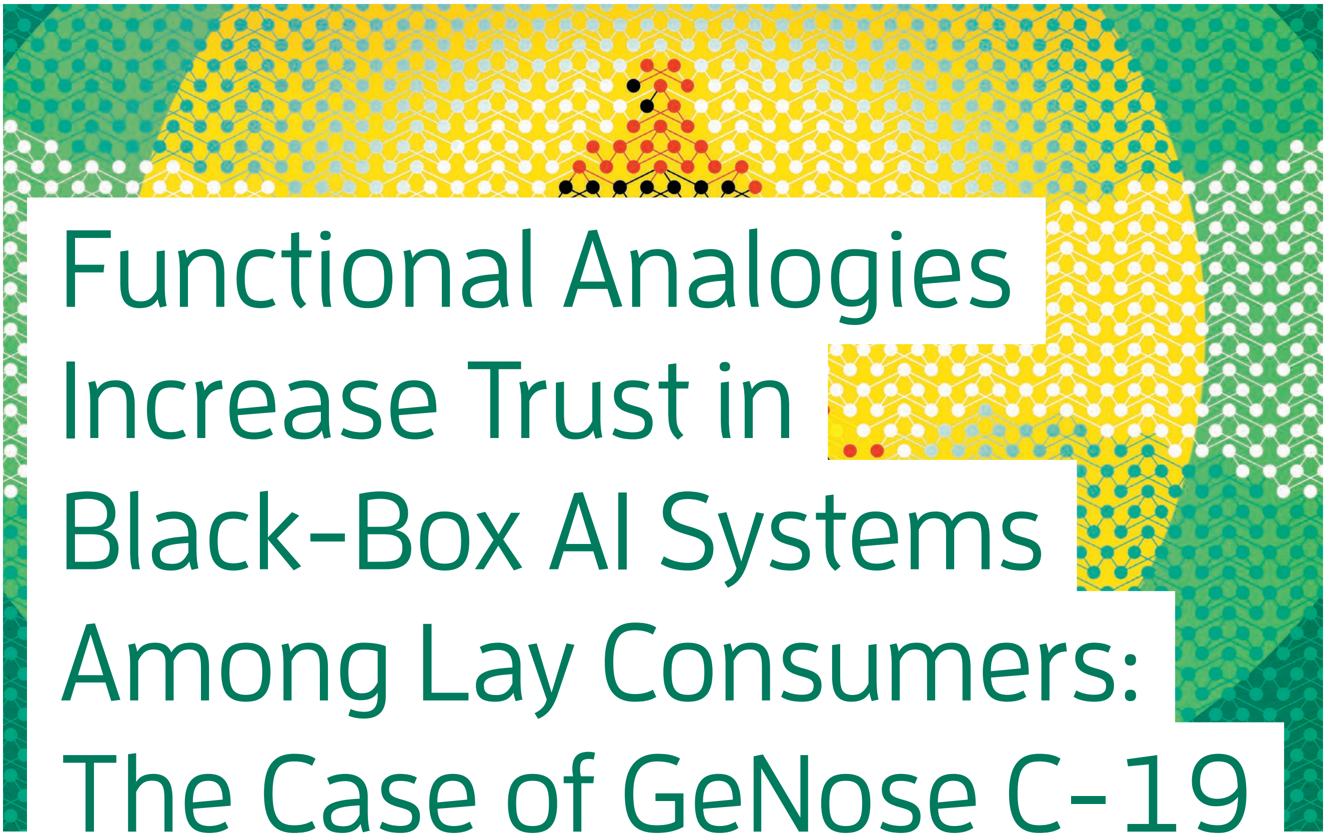

Functional Analogies Increase Trust in Black-Box AI Systems Among Lay Consumers: The Case of GeNose C-19

Elizabeth Demissie Degefe, Duke University
Yulius Denny Prabowo, Institut Teknologi dan Bisnis Kalbis
Krishna Savani, The Hong Kong Polytechnic University
Abhishek Sheetal, The Hong Kong Polytechnic University and Central Queensland University

We present a case study of a black-box artificial intelligence-based COVID-19 detection product, GeNose C-19, developed by the Indonesian government. We find that explaining how GeNose works using functional analogies increases both Indonesian and American lay consumers' trust in GeNose.

Artificial intelligence (AI)-based products often succeed in the market but also fail. Similar to traditional technological products, some market failures are due to technological shortcomings, whereas others are due to overpromises made via marketing that the technology could not deliver.[1] However, some products fail to get market traction despite their cutting-edge and functional technology simply because the general public distrusts them. An example of such a market failure is GeNose C-19 (hereafter GeNose), an AI-based COVID-19 detection tool.[2]

GeNose was developed by researchers at Universitas Gadjah Mada, Indonesia. It was built on the research concept of an electronic nose, which was later productized as a COVID-19 detection system. An electronic nose is a sensor that converts airborne particles into a sequence of numbers. Although these numbers have no meaning by themselves (similar to the pixels of an image), at a higher level of abstraction, they carry the fingerprints of

Digital Object Identifier 10.1109/MC.2023.3235880
Date of current version: 3 May 2023

 0018-9162/23©2023IEEE

a specific smell. GeNose uses an electronic nose to detect the distinctive pattern of volatile organic compounds in the breath of people infected with COVID-19.[3] A stream of numbers representing various uninterpretable aspects of smell cannot be meaningfully modeled using traditional regressions; it would be akin to running regressions on the pixels of an image. Instead, the developers trained a machine learning model to predict the outcome of individuals' COVID-19 polymerase chain reaction (PCR) tests based on the numeric fingerprints of their breath. The training input data comprised the numeric data generated by the electronic nose along with participants' demographics, such as age, gender, and smoking status. The training output data were positive/negative flags from the PCR test.

Once the model was trained, the developers tested the GeNose product in diagnostic trials across seven hospitals. GeNose was able to predict whether people had COVID-19 (based on the results of a PCR test) with a sensitivity of 89%–92% and specificity of 95%, rivaling that of the antigen rapid test (ART; 76.3% sensitivity and 98.9% specificity) and approaching that of the PCR test (90.7% sensitivity and 100% specificity).[4,5,6] In addition, GeNose could detect viral infections at an earlier time point than ART and PCR tests (two days after infection instead of four days), offered results within 3 min, and was 91%–99% cheaper than PCR and ART tests (see Table 1). Travelers crossing checkpoints such as train stations, airports, and state boundaries merely had to queue up at the GeNose testing site, blow into a plastic bag, and get a clearance to cross the zone within minutes.

Although GeNose is cheap and reliable; safe, easy, and convenient to use; transparent in terms of how it was developed (i.e., based on open-sourced product architecture and algorithms); and built by researchers at an educational institution as a public good (rather than for profit making), the general public did not react well to GeNose.[7] Initially, the Indonesian government perceived GeNose as a viable option for screening domestic and international travelers for COVID-19.[2] To limit the spread of COVID-19, the Indonesian government required all domestic and international passengers to be screened for COVID-19 before they traveled by airplane, train, or bus.[8] The fourth most populous country in the world and a developing country, Indonesia needed a product that was affordable and could rapidly detect COVID-19 at its many bus stations, train stations, and airports, a task that GeNose was perfectly suited for.[2] However, weeks after its deployment, the general public started to question GeNose's effectiveness, with Indonesians selectively focusing on GeNose's misdiagnoses.[7] People demanded that the government halt GeNose screenings at train stations.[9]

The government's decision to stop using GeNose to screen tourists appeared to confirm the public's distrust.[10] The discourse associated with GeNose highlighted that aside from cases of inaccuracy, people distrusted GeNose because they did not understand how it worked, with some doubting the existence of breath-based evidence of COVID-19 infection.[11] Although inaccuracy is certainly a concern, even ART and PCR tests can be inaccurate. Yet there has been no discourse about the inaccuracy of these tests, likely because they are based on old technologies. However, as GeNose is based on new AI technology, laypeople questioned its inner working.

The general public often distrusts AI technology as people are apprehensive of products that they do not understand.[12] This confusion likely arises because people cannot understand how AI systems make decisions. Without such an understanding, people have difficulty trusting the reliability and accuracy of AI systems. Thus, giving people an intuitive sense of how AI systems work is a vital prerequisite for people to trust AI-based products.[13] For example, within human–robot teams, a lack of transparency about how robotic teammates make their predictions led human participants to distrust their robotic partner's decisions.[14] People might even perceive decisions made by AI systems

TABLE 1. Comparison of tests based on data acquired from past research.[4,5,6]

Test Method	Detectable Level After Viral Exposure	Time to Test Results	Sensitivity (%)	Specificity (%)	Cost per Test (US$)
RT-PCR	Four days	12 h to one week	90.7	100	63–247
ART	Four days	15–30 min	76.3	98.9	19.4
GeNose C-19	Two days	3 min	89–92	95	0.7–1.7

as random as they do not see any consistent set of rules by which such a system functions.[15] We propose that a similar phenomenon is occurring with GeNose.

Much of the current work on transparency and trust in AI is centered around populations who have a background in technology, such as engineers and researchers, thus making transparency an issue examined only from the developer's and regulator's perspective. However, many AI products are marketed to laypeople who have little to no understanding of the underlying technology, and thus, the solutions that increase experts' trust in AI are likely to differ from those that increase laypeople's trust. Instead of understanding the fine details of the underlying machine learning algorithms, laypeople just need an intuitive understanding of how the product works.[16] In the case of GeNose, the public does not understand how their breath is relevant to their COVID-19 diagnosis and how GeNose's AI algorithm can detect COVID-19 from their breath, and thus, they do not trust GeNose. Hence, the case of GeNose highlighted an often ignored dimension of trustworthiness—the perspective of laypeople.

Although it is possible to create a marketing plan for a transparent AI product,[17] it is particularly challenging with black-box AI systems, such as GeNose, as the systems' processes are not meant to be apparent. To counter this inability to explain how the AI system functions, marketers have often focused on highlighting how AI systems can counter human biases and can be used to meet human objectives, thus improving consumers' trust in the product.[16] However, marketers can also build on research on functional analogies to increase the public's understanding of black-box AI systems.[18] Functional analogies explain how a new product works by building on consumers' existing knowledge.[19] Analogies allow consumers to identify similarities between the concepts they are already familiar with and the new product, thus gaining some understanding of how the new product works. The process of analogical learning encompasses three stages: access, mapping, and inference.[18]

In the first stage of analogical learning, access, common everyday products, or experiences are highlighted to activate consumers' preexisting knowledge.[18] These common products or experiences then act as a base that consumers can use as a reference point in relation to the new product. With an established base, consumers then enter the mapping phase in which they identify similarities between components of the existing product and the new product.[18] That is, consumers identify how components of the base product and the new product share similar roles in how the two products function. For example, consumers might perceive that the two products use similar processes even if they are physically different. After identifying a common relational structure, consumers then enter the inference stage in which they transfer their knowledge of the base product to understand how the new product functions.[18] This greater understanding of how the new product works then improves consumers' perceptions of the product's transparency, and consequently, their trust in the new product.

However, to activate this analogical learning process, the analogy between the new product and the preexisting product must be sufficiently compelling. The preexisting product cannot simply share superficial similarities with the new product; it must do so at a deep underlying level as consumers view superficial similarities as inappropriate.[19] If analogies are perceived as unsound or irrelevant, consumers do not engage in any inference or knowledge transfer. Thus, analogies need to be factually accurate with reference to how the two products work to engender any understanding of the new product. In addition, analogous products should also be carefully selected in terms of the emotions that they evoke in consumers as the inference stage of analogical learning not only transfers knowledge but also emotional connotations.[19] Thus, analogies should have the same emotional associations that developers wish consumers to associate with the new product, a consideration that requires analogies to be culturally appropriate as emotional associations vary across cultures.

The current study seeks to use functional analogies to explain unsupervised machine learning models and make them more transparent to laypeople across two markets. Specifically, we hypothesize that the use of functional analogies will increase laypeople's trust in GeNose. Data for the studies are shared at https://doi.org/10.17605/osf.io/gjn6h.

METHOD

Study 1

Study approach and design. We first assessed the effect of our analogy-based explanation in Indonesia. Specifically, we tested whether our analogies could improve the Indonesian public's trust in the product. To create our functional analogies and demystify how the AI product works, we first assessed which aspects of GeNose laypeople would have difficulty understanding. We determined that the aspects of GeNose that needed greater transparency were 1) how COVID-19 could be evident within the breath and 2) how the GeNose machine could detect this breath-based evidence. We sought to explain both these features by drawing parallels with people's everyday experiences.

We identified everyday experiences that would serve as effective functional analogies by examining how they met two criteria.[19] First, the analogous experience needed to be sound, such that it accurately addressed and represented the functioning of the product. Thus, the structure of the analogous experience needed to closely parallel that of the product. Second, the analogy needed to contain information that the majority of laypeople would already be familiar with, thus ensuring that participants have the reference points needed to map the similarities between the analogy and the product.

To explain how COVID-19 can be detected from the breath, we compare patterns of volatile organic compounds in people's breath to two different strong scents that people can often detect on others' breath—coffee and jamu, an Indonesian herbal medicine. The smell of coffee is considered pleasant in Indonesia and is a common scent in Indonesian cosmetic products. Jamu, on the other hand, is a traditional herbal drink consumed for health benefits that has an unpleasant smell. When people drink coffee or jamu, their breath smells a bit like coffee or jamu, at least for some time. This way, we connect the abstract concept of volatile compounds in breath to respondents' everyday experiences with different strong scents of the breath: an analogy that meets both of our criteria for soundness and familiarity.

To explain how GeNose can detect COVID-19 from people's breath, we invoke the concept of fingerprints. GeNose uses the unique concentrations of various volatile organic compounds in breath samples to identify COVID-19. Similarly, individuals' unique fingerprints are used to identify them in everyday life. This analogy is sound because both fingerprints and organic compounds in breath samples are biological patterns used for the purpose of identification. The analogy is also familiar as most laypeople are aware of fingerprints and their role in identifying individuals. We then used these functional analogies to create a short essay introducing GeNose and explaining how it works.

TO EXPLAIN HOW GENOSE CAN DETECT COVID-19 FROM PEOPLE'S BREATH, WE INVOKE THE CONCEPT OF FINGERPRINTS.

User study procedure. Our study was approved by the Institutional Review Board at Nanyang Technological University and the Institut Teknologi dan Bisnis Kalbis. After they provided informed consent, respondents were randomly assigned to either the analogy-based or technical explanatory essays of GeNose. The analogy-based essay read as follows:

> "Human breath contains thousands of microparticles that create different smells depending on their concentration. Each smell is made up of specific particles that create a fingerprint in your breath. While humans cannot sense most of these particles, they are easily detectable. Just like dogs are more sensitive to smell than humans, machines are more sensitive than any living organism because they can detect these microparticles even in very small concentrations.
>
> Using highly sensitive olfaction detection machines, scientists created GeNose C-19, a device that can smell the Covid-19 fingerprint on your breath. As an analogy, imagine your breath can have two distinct smells, both of which are too subtle for the human nose to detect.
>
> If you are healthy, it smells a bit like coffee. If you have Covid-19, it smells a bit like jamu.
>
> When you breathe into our machine, it catches these microparticles and determines if you have the jamu fingerprint or the coffee fingerprint.
>
> Using advanced AI technology, scientists discovered Covid-19's unique fingerprint. GeNose C-19 scans your breath for this Covid-19 fingerprint to determine if you are infected. This innovative technology can detect covid with 95% accuracy, rivaling that of other Covid tests such as the antigen rapid test (ART)."

In the control condition, we presented participants with a parallel essay based on the original GeNose press release, which was meant to inform the general public about the product.[6]

> "To help people better test for Covid-19, scientists created GeNose C-19, a breath-based COVID-19 detection device. GeNose C-19 is an electronic nose that mimics the mechanism of the human nose to detect Covid-19. Using a machine-learning-based AI system, it was trained to recognize the distinctive response pattern of people infected with Covid-19.
>
> A team of researchers examined the Volatile Organic Compounds (VOCs) exhaled by Covid-19 patients. VOC is composed of various gases exhaled when we breathe. These gasses are formed by many biological mechanisms such as inflammation caused by viral or bacterial infections. Scientists compared the composition of expiratory gases exhaled by 316 COVID-19 patients to 299 patients without COVID-19. They discovered a unique pattern in the composition of VOCs exhaled by those infected with Covid-19.
>
> This unique VOC pattern in Covid-19 patients was not affected by existing conditions, such as hypertension or asthma. Moreover, it was distinct from VOC patterns that occur from infections of other viruses, such as the common cold and flu, and bacterial infections. GeNose C-19 is capable of detecting the Covid-19 VOC pattern in your breath and can determine if you have Covid with 95% accuracy, rivaling that of other Covid tests such as the antigen rapid test (ART)."

Note that such an essay would likely be optimal for addressing issues of transparency on the developer side as it contains the technical details those with technical experience would focus on.

After participants read the respective articles, we measured the extent to which they trusted GeNose by adapting an existing product trust measure[20] (α = 0.92). Our scale had two dimensions. The first dimension asked participants if they would personally use and trust GeNose.

1. "Would you trust or distrust that GeNose C-19 can accurately detect COVID-19?"
2. "Do you believe that GeNose C-19 can actually detect COVID-19 with the accuracy that it claims?"
3. "If you had to choose between GeNose C-19 and ART tests to find out if you have COVID-19, which would you choose?" (a = 0.83.)

The second dimension asked whether participants would advocate for GeNose to be used within public spaces.

4. "Do you think restaurants should use GeNose C-19 to test whether their customers have COVID-19?"
5. "Do you think offices should use GeNose or ART tests to find out if employees have COVID-19?"
6. "Do you think the government should use GeNose or ART tests to monitor the spread of COVID-19 in the population?" (α = 0.83.)

For questions 1, 2, and 5, we used a six-point response scale ranging from one ("strongly distrust/strongly disbelieve/definitely should not") to six ("strongly trust/strongly believe/definitely should") For the remaining questions, we used a six-point response scale ranging from one ("definitely ART") to six ("definitely GeNose"). All measures were translated to Bahasa Indonesia, the official language of Indonesia.

Respondents. To recruit Indonesian respondents, we used an online recruitment platform that specializes in recruitment within the Asian market, dataSpring. We sought participants who were at least 18 years of age. We recruited 149 adults (62 women, 85 men, and two identifying as other genders) with an average age of 34.03 years and a standard deviation (SD) of 8.79 years. Our sample came from a variety of Indonesian ethnic groups, with approximately 59% identifying as Javanese, 26% as Sundanese, 4% as Chinese, 1% as Betawi, 1% as Madurese, 5% as other race, and 5% choosing not to disclose their ethnic background.

Results. An independent sample t-test found that participants had higher trust in GeNose in the analogy condition [mean = 4.04; SD = 1.01; 95% confidence interval (CI) (3.81, 4.27)] than in the control condition [mean = 3.73; SD = 1.07; 95% CI (3.49, 3.98); p = 0.036 (one-tailed); Student's t-test t(147) = 1.82; see Figure 1]. The effect size was small to medium; Cohen's d = 0.30; 95% CI (–0.03, 0.62). We found similar effects for both dimensions included in the scale (see Table 2).

Study 2

The relatively low effect size in Study 1 is likely because many Indonesians are already familiar with GeNose and have entrenched attitudes about the product that are difficult to shift. Functional analogies may be most effective in increasing trust in new AI-based technologies when a product is first introduced to a market. Thus, we next sought to test our functional analogy-based explanation in the United States, a market unfamiliar with GeNose.

User study procedure. The experimental procedure was identical to that used in Study 1 with one minor change. To ensure that our analogy was culturally appropriate, we compared the breath of people with COVID-19 to garlic instead of jamu. Similar to jamu in Indonesia, the smell of garlic is typically considered unpleasant in the United States. We used the same measure to assess trust in GeNose as in Study 1 ($\alpha = 0.90$), assessing both personal trust ($\alpha = 0.85$) and advocacy for public use ($\alpha = 0.83$). The survey was conducted in English.

Respondents. To recruit U.S. respondents, we used the Amazon Mechanical Turk online recruitment platform. As a precaution for data quality, we also recruited only participants who were marked as being high-quality participants by CloudResearch, which samples only individuals who have received high approval ratings in previous surveys. We sought participants who were at least 18 years of age. We recruited 92 adults (45 women and 47 men) with a mean age of 41.23 years and an SD of 11.15 years. Our sample was ethnically diverse; approximately 73% of the participants identified as European Americans, 13% as African American, 6% as Asian American, 2% as Latinx, 2% as Native American, and 4% as mixed race.

Results. An independent samples t-test indicated that participants in the analogy condition [mean = 4.48; SD = 1.04;

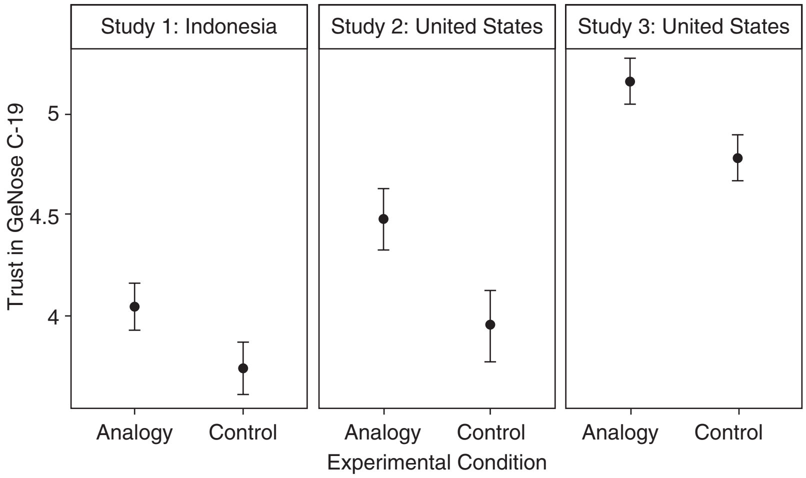

FIGURE 1. Trust in GeNose C-19 across studies. Error bars represent the standard error of the mean.

TABLE 2. Trust in GeNose C-19 across studies, broken down by trust subdimensions. Sample size: $N = 149$ in Study 1, $N = 92$ in Study 2, and $N = 150$ in Study 3.

Study	Dimension	Mean (SD)		p	Cohen's d
		Analogy	Control		
1	Personal trust	4.15 (.95)	3.88 (1.03)	0.04	0.28
1	Public advocacy	3.93 (1.15)	3.59 (1.18)	0.04	0.29
2	Personal trust	4.83 (1.13)	4.36 (1.30)	0.03	0.39
2	Public advocacy	4.12 (1.09)	3.53 (1.35)	0.01	0.49
3	Reliability	5.41 (1.37)	5.13 (1.34)	0.10	0.21
3	Maintainability	5.51 (1.27)	5.03 (1.21)	0.01	0.39
3	Safety	5.36 (1.38)	4.96 (1.37)	0.04	0.29
3	Security	4.55 (1.71)	4.15 (1.52)	0.07	0.25
3	Privacy	4.45 (1.65)	4.03 (1.37)	0.04	0.28
3	Usability	5.71 (1.25)	5.24 (1.30)	0.01	0.37
3	Ethics	5.13 (1.20)	4.71 (1.28)	0.02	0.34
3	Robustness	5.17 (1.35)	5.03 (1.34)	0.25	0.11

95% CI (4.17, 4.78)] trusted GeNose significantly more than those in the control condition [mean = 3.94; SD = 1.25; 95% CI (3.59, 4.30); $p = 0.015$ (one-tailed); t(90) = 2.21; Cohen's d = 0.46; 95% CI (0.04, 0.88)]. The effect size was about 53% larger in the United States than in Indonesia, probably because Americans did not have preformed attitudes about GeNose and were thus more influenced by the functional analogies. We found similar effects for the two dimensions of trust (see Table 2).

Study 3

Although Studies 1 and 2 provided support for our key hypothesis, they used general measures of trust. However, trust is a multidimensional construct with many subcomponents. Study 3 thus measured numerous dimensions of trust: reliability, maintainability, safety, security, privacy, usability, ethics, and robustness. We hypothesized that functional analogies would increase people's trust in all these dimensions because an understanding of the product is needed for people to comprehensively trust the product.

User study procedure. The experimental procedure was identical to that used in Study 2 with the exception of a new measure of trust. In this measure, each item assessed a different dimension of trust.

1. "GeNose C19 can reliably detect COVID-19 better than other COVID-19 tests" (reliability).
2. "It will be easy for businesses to maintain GeNose C19 as a tool to detect COVID-19" (maintainability).
3. "GeNose C19 is a safer way to test for COVID-19 than other tests such as ARTs" (safety).
4. "The data collected from GeNose C19 poses less of a threat to my security than that collected by other COVID-19 tests such as ARTs" (security).
5. "The data collected from GeNose C19 poses less of a threat to my privacy than that collected by other COVID-19 tests such as ARTs" (privacy).
6. "GeNose C19 is easier to use than other tests such as ARTs" (usability).
7. "GeNose C19 is an ethically made product" (ethics).
8. "GeNose C19 provides more consistent results than other COVID-19 tests such as ARTs" (robustness) ($\alpha = 0.88$).

All the items were assessed on a seven-point response scale ranging from one ("strongly disagree") to seven ("strongly agree"). The study was conducted in English.

Respondents. We recruited U.S. respondents using the Amazon Mechanical Turk online recruitment platform. Again, we recruited only those above 18 years of age and those marked as high-quality participants by CloudResearch. We recruited 150 adults (70 women and 80 men) with a mean age of 37.4 years and an SD of 11.01 years. The ethnic composition of our sample was as follows: 80% of the participants identified as European Americans, 8% as African American, 5% as Asian American, 2% as Latinx, 1% as Native American, 3% as mixed race, and 1% as other.

Results. We first examined the overall trust measure averaging across all items. An independent samples t-test found that participants had higher trust in GeNose in the analogy condition [mean = 5.16; SD = 1.03; 95% CI (4.92, 5.39)] than in the control condition [mean = 4.78; SD = 0.96; 95% CI (4.57, 5.00); $p = 0.010$ (one-tailed); t(148) = 2.33; Cohen's d = 0.38; 95% CI (0.06, 0.71)]. Upon examining the various dimensions of trust, we found that participants in the analogy condition displayed higher levels of trust in GeNose across multiple dimensions, including maintainability, safety, privacy, usability, and ethics (see Table 2). We found nonsignificant effects for reliability, security, and robustness, indicating that increasing trust on these dimensions possibly requires more than analogies. Overall, this study indicates that functional analogies increase trust across multiple dimensions.

DISCUSSION

New AI-based technologies and products are useless if the intended consumers do not accept them once they are launched. This research identified an easy-to-implement method to increase people's trust in new AI-based products and technologies—using functional analogies to give laypeople a rough sense of how the product works. People are often skeptical of new AI-based products because they have no idea how the products function. However, analogies help people understand how the product works by referring to phenomena that people already understand in their everyday lives, thereby demystifying the product. If people have some idea of how the product works, they are more likely to trust it. Our work furthers the issue of trustworthiness in AI by addressing how one of the vital prerequisites of trust, transparency, can be improved within a lay population that has little background in technology. We found that functional analogies are an effective way to improve laypeople's understanding of AI systems.

The functional analogies used need not be exact—they only need to give people a rough sense of how the product functions. Using this procedure, we increased people's trust in GeNose, an AI-based COVID-19 detection tool, both in the home country where the product was developed and launched and in a new country where the product had not yet been introduced. The effect of functional analogies was stronger in the country where people were unfamiliar with GeNose, indicating that functional analogies are particularly beneficial when AI-based products are first introduced to a market. Functional analogies were effective even if people had entrenched attitudes, but their benefit was smaller in magnitude. We also found that functional analogies were able to improve trust in GeNose even when using a measure that assessed multiple dimensions of trust and were, in fact, more effective in improving the practical aspects of trust, such as usability, maintainability, and safety, than more technical aspects, such as robustness and reliability.

This focus on practical aspects is likely because laypeople have limited technical understanding and so are unlikely to differentiate between products based on trust in their technical ability. Instead, as their primary concern is deciding whether to use a product, they would have a greater focus on whether products can be trusted in practice. This effect indicates a possible difference in the way that layperson and expert trust is developed, meaning that different methods may be needed to improve each. Overall, these findings suggest that popular press releases and advertising campaigns for new AI-based technologies and products should focus more on using analogies to give people a sense of how the product works in addition to detailing the technical aspects of the product to improve trust in the product's practical abilities.

This approach to improving AI trust is relevant to all domains in which complex technologies need to be judged by the general public. Functional analogies allow the average consumer with no background in the field to understand how a complex product functions with fairly little difficulty. It is consumers' difficulty in understanding products that creates a sense of distrust, a refusal to use the product, and the eventual market failure of the product. Any complex technology can avoid this fate by using functional analogies that can help improve laypeople's understanding of the product, and therefore, increase their trust in the product.

We used a random analogy for part of our analogy-based explanation in our current experiment as this study was supposed to provide a proof of concept. The breath of people with COVID-19 probably does not smell like coffee, garlic, or jamu. The analogy could have referenced any pleasant or unpleasant scent that people experience daily to create a similar effect. However, analogies need to be sound and familiar to gain the public's trust in the long run. To do this, developers can train machine learning models to identify factually accurate yet culturally appropriate fingerprints to use as analogies. This way, the analogies used can be defended if challenged. C

ACKNOWLEDGEMENT

We thank Whiwdon Jagad Uki and Altobelli Lobodally for their help in the design and execution of experiments in Indonesia. Parts of this research were conducted when Elizabeth Degefe, Krishna Savani, and Abhishek Sheetal were at Nanyang Technological University. We acknowledge Nanyang Technological University's support in funding this research. This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Boards at Nanyang Technological University under protocol IRB-2015-07-018 and at the Institut Teknologi dan Bisnis Kalbis under protocol titled "Studying the perceptions of AI in society," and performed in line with the Declaration of Helsinki. The corresponding author is Abhishek Sheetal.

REFERENCES

1. T. R. Weiss, "IBM Watson health finally sold by IBM after 11 months of rumors," *EnterpriseAI*, 2022. Accessed: Jan. 25, 2022. [Online]. Available: https://www.enterpriseai.news/2022/01/21/ibm-watson-health-finally-sold-by-ibm-after-11-months-of-rumors
2. M. Taylor, "Indonesia to introduce new GeNose COVID-19 test for foreign visitors to Bali," *Thaiger*, 2021. Accessed: Jan. 25, 2022. [Online]. Available: https://thethaiger.com/news/regional/indonesiaen/indonesia-to-introduce-new-genose-covid-19-test-for-foreign-visitors-to-bali
3. D. Nurputra, "Genosvid diagnostic test for early detection of COVID-19," U.S. Nat. Library Med., Bethesda, MD, USA, Sep. 2020. Accessed: Feb. 11, 2022. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04558372?cond=Covid19&cntry=ID&city=Yogyakarta&draw=2
4. L. E. Brümmer et al., "Accuracy of novel antigen rapid diagnostics for

ABOUT THE AUTHORS

ELIZABETH DEMISSIE DEGEFE is a postdoctoral associate at the Samuel Dubois Cook Center and the Fuqua School of Business, Duke University, Durham, NC 27708-0120 USA. Her research interests include using machine learning methods to understand culture and diversity in organizations. Degefe received a Ph.D. in organizational behavior from Nanyang Technological University, Singapore, and was at Nanyang Technological University when this research was conducted. Contact her at elizabeth.degefe@duke.edu.

YULIUS DENNY PRABOWO is a senior lecturer at the Informatics Department of Institut Teknologi dan Bisnis Kalbis, East Jakarta 13210, Indonesia. His research interests include natural language processing, particularly in the Indonesian languages, and using text summarization to analyze publicly available data and articles to create an Indonesian text summarization corpus. Prabowo received a doctorate in computer science from Bina Nusantara University. Contact him at yuliusdennypr@gmail.com.

KRISHNA SAVANI is a professor at The Hong Kong Polytechnic University, Hong Kong, China. His research interests include studying culture, diversity, and decision making and using machine learning to generate hypotheses and make discoveries in the social sciences. Savani received a Ph.D. in psychology from Stanford University and was at Nanyang Technological University when this research was conducted. Contact him at krishna.savani@polyu.edu.hk.

ABHISHEK SHEETAL is the managing director of the Centre for Leadership and Innovation at The Hong Kong Polytechnic University, Hong Kong, China, where he works on applied artificial intelligence (AI) research. His research interests include using AI to solve everyday societal problems and to foster greater adoption of machine learning methods in social sciences and public policy research. Sheetal received an M.B.A. from Columbia Business School. Contact him at asheetal12@gsb.columbia.edu.

SARS-COV-2: A living systematic review and meta-analysis," *PLoS Med.*, vol. 18, no. 8, Aug. 2021, Art. no. e1003735, doi: 10.1371/journal.pmed.1003735.
5. J. N. Kanji et al., "False negative rate of COVID-19 PCR testing: A discordant testing analysis," *Virology J.*, vol. 18, no. 1, pp. 1–6, Jan. 2021, doi: 10.1186/s12985-021-01489-0.
6. D. K. Nurputra, K. Triyana, and T. H. Sasongko, "Indonesian-made COVID-19 breathalyser sensitivity comparable to RTPCR," The Conversation, Carlton, Australia, 2021. Accessed: Feb. 11, 2022. [Online]. Available: https://theconversation.com/indonesian-made-covid-19-breathalysersensitivity-comparable-to-rt-pcr-155497
7. F. Rosana, "YLKI reveals cases of positive GeNose tests, but negative during retest with PCR," *Tempo*, 2021. Accessed: Feb. 11, 2022. [Online]. Available: https://bisnis.tempo.co/read/1447824/ylki-ungkap-kasus-tes-genose-positif-tapinegatif-saat-tes-ulang-dengan-pcr
8. "COVID-19 task force issues new regulation for domestic travel during pandemic," Cabinet Secretariat of the Republic of Indonesia, Central Jakarta, Indonesia, 2021. Accessed: Feb. 11, 2022. [Online]. Available: https://setkab.go.id/en/covid-19-task-force-issues-new-regulation-for-domestic-travel-during-pandemic/
9. R. Haryanti, "GeNose begins to be used at stations, residents: Cheap, but effective?" *Kompas*, Feb. 2021. Accessed: Feb. 11, 2022. [Online]. Available: https://megapolitan.kompas.com/read/2021/02/03/21154221/genose-mulai-digunakan-di-stasiun-warga-murah-sih-tapi-efektif
10. D. Praptono, "GeNose C-19 test is no longer valid when entering Bali because it is not accurate," *Radar Bali*, 2021. Accessed: Feb. 11, 2022. [Online]. Available: https://radarbali.jawapos.com/berita-daerah/dwipa/28/06/2021/tes-genose-c-19-tak-berlaku-lagi-saat-masuk-bali-karena-tak-akurat
11. A. Zuhad, "How it works and disadvantages of GeNose, a COVID-19 detection tool made by UGM," *Kompas*, 2021. Accessed: Feb. 24, 2022. [Online]. Available: https://www.kompas.tv/article/141050/cara-kerja-dan-kekurangangenose-alat-deteksi-covid-19-buatan-ugm?page=all
12. T. Ye et al., "Psychosocial factors affecting artificial intelligence adoption in health care in China: Cross-sectional study," *J. Med. Internet Res.*, vol. 21, no. 10, Oct. 2019, Art. no. e14316, doi: 10.2196/14316. [Online]. Available: http://www.jmir.org/2019/10/e14316/
13. E. Glikson and A. Williams, "Human trust in artificial intelligence: Review of empirical research," *Acad. Manage. Ann.*, vol. 14, no. 2, pp. 627–660, Jul. 2020, doi: 10.5465/annals.2018.0057.

14. J. Chen and M. Barnes, "Human—agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014, doi: 10.1109/THMS.2013.2293535.
15. K. Stubbs, P. Hinds, and D. Wettergreen, "Autonomy and common ground in human-robot interaction: A field study," *IEEE Intell. Syst.*, vol. 22, no. 2, pp. 42–50, Mar./Apr. 2007, doi: 10.1109/MIS.2007.21.
16. A. Rai, "Explainable AI: From black box to glass box," *J. Acad. Marketing Sci.*, vol. 48, no. 1, pp. 137–141, Jan. 2020, doi: 10.1007/s11747-019-00710-5.
17. A. Wind, E. Constantinides, and S. de Vries, "Marketing a transparent Artificial Intelligence (AI): A preliminary study on message design," DANS, The Hague, The Netherlands, 2019. Accessed: Jan. 25, 2022. [Online]. Available: https://www.narcis.nl/publication/RecordID/oai:ris.utwente.nl:publications%2F2db90e0a-0935-4504-b19d-c1ccefd5b428
18. J. Gregan-Paxton, J. D. Hibbard, F. F. Brunel, and P. Azar, "'So that's what that is': Examining the impact of analogy on consumers' knowledge development for really new products," *Psychol. Marketing*, vol. 19, no. 6, pp. 533–550, Jun. 2002, doi: 10.1002/mar.10023.
19. M. R. Goode, D. W. Dahl, and C. P. Moreau, "The effect of experiential analogies on consumer perceptions and attitudes," *J. Marketing Res.*, vol. 47, no. 2, pp. 274–286, Apr. 2010, doi: 10.1509/jmkr.47.2.274.
20. E. P. Becerra and P. K. Korgaonkar, "Effects of trust beliefs on consumers' online intentions," *Eur. J. Marketing*, vol. 45, no. 6, pp. 936–962, May 2011, doi: 10.1108/03090561111119921.

Digital Object Identifier 10.1109/MC.2023.3263911