# Trust and transparency in privacy-enhancing technologies

Privacy Enhancing Technologies

ODI Research
ADVANCING TRUST IN DATA

February 2025

# Executive summary

Privacy-enhancing technologies (PETs) can provide organisations with enhanced security and confidentiality of data and code, but several challenges and barriers stand in the way of their adoption. One such barrier is a lack of trust: sceptical, privacy-aware individuals find it difficult to trust that PETs will effectively keep their sensitive data and code protected and safe.

In this research, we sought to understand trust in PETs and the factors that motivate and dissuade their adoption. We aimed to analyse different transparency measures that PET providers can use to increase trust, and the extent to which they do so in practice.

We focused on the context of trusted execution environments (TEEs), a type of PET with undeniable benefits for organisations looking to use confidential computing. TEEs also present particularly interesting dynamics around their trust and adoption, and we used Google's Project Oak, an in-development piece of TEE infrastructure designed with a number of transparency measures, as a framework to explore them.

Using qualitative data collected from three activities, we employed thematic analysis to determine: (i) transparency measures only motivate trust when they are meaningful; (ii) when it comes to trust in PETs, principles often trump efficacy; (iii) Google's Project Oak's transparency measures have strengths and limitations; and (iv) technical transparency measures only address certain concerns for certain actors.

We found that the onus on PET providers to make trust decisions easier, which requires both transparency and the use of socio-technical measures, was an overarching theme. We briefly explored how these findings can be broadened to other types of PETs and proposed a set of recommendations for further research and future work.

# Contents

# About

This report has been researched and produced by the Open Data Institute (ODI), and published in February 2025. Its authors were Neil Majithia, Calum Inverarity and Elena Simperl with support from Ruba Abu-Salma and Claudine Tinsman. If you want to share feedback by email or would like to get in touch, contact the privacy-enhancing technologies (PETs) programme at pets@theodi.org.

This report has been made possible through the support provided by the team at Google Research.

# Introduction

Privacy-enhancing technologies (PETs) are tools and practices that can enable the access to, and the usage of, data that might otherwise be kept closed due to privacy concerns, like individual medical records or commercially sensitive information.[1] By doing so, PETs can enable organisations to perform analyses that would be difficult or impossible if the data were to be kept closed.

In previous research, we have considered the barriers to the adoption of PETs, such as the lack of knowledge and resources that explain how these technologies work in practice.[2] Given the sensitivity of the data being collated, accessed and shared, it is understandable that potential adopters unfamiliar with the technologies might question the veracity of PET providers' privacy and security claims.

In spite of the premise of increased privacy and security that PETs bring, and the potential benefits of their usage, the decision to adopt a PET and incorporate it into an organisation's technology stack is not simple. A potential adopter must consider the resource costs and technical expertise required to adopt some PETs. Some PETs must be procured from external organisations (hereafter referred to as 'providers'), which requires trust between the two parties in the exchange. Both of these constraints are extremely relevant in the case of one specific type of PET – trusted execution environment (TEE) architectures.

TEEs are the core components of confidential computing, a powerful privacy-enhancing measure that an organisation can use to run its code and analyse its data on an untrusted device. Applications of confidential computing include analysis of sensitive data, such as user health or financial data.[3] However, the adoption and usage of TEEs requires high levels of confidence in the privacy and security guarantees given by the software and hardware – both of which are developed and maintained by a third-party organisation – to the adopting organisation. Simply put, for the TEE to be adopted and used in the most sensitive of cases, the adopter must trust the technology and its providers. For TEEs, this will likely include a variety of actors, including cloud hosts, providers of any APIs used, and, arguably more fundamentally, chip manufacturers.

---

[1] ODI (2023), 'Privacy enhancing technologies (PETs)'.
[2] ODI (2024,) 'PETs in Practice 1', 'PETs in Practice 2' and 'PETs in Practice 3'.
[3] Geppert, T., Deml, S., Sturzenegger, D., and Ebert, N. (2022), 'Trusted Execution Environments: Applications and Organizational Challenges'.

On this basis, our research considered how trust in PETs is motivated and dissuaded. Specifically, we sought to examine how a provider can incorporate certain transparency features into their PET to help build trust, and the limitations to this approach. To do so, we focused on the context of TEEs and, specifically, Google's Project Oak.

# Background

The ODI has previously carried out research on the importance of trust and how this affects data sharing between different actors, including both individuals and organisations. Examples include our research on who people in the UK, Germany, France, Belgium and the Netherlands trust with personal data[4] and our broader programme of work on data assurance.[5] Through this research, we have considered the data practices and behaviours that organisations should adhere to in order to build and maintain the trust of the people they serve. These practices have primarily been grounded in responsible data governance. However, PETs provide additional means through which organisations can provide further privacy guarantees that can contribute towards greater trust, and therefore greater availability of data.

The relationship between trust and transparency in emerging technologies, however, is complex and under-analysed. A simple assumption would posit that increased transparency can help generate greater trust. However, this fails to account for many contributing factors (eg, an individual or organisation's general views on technology, professional background or sector, technical literacy and expertise, demographics, regulations). These factors may all influence the level of transparency required to achieve a particular objective, for example product adoption or increased user trust. Research in human-centric PETs has shown that people's opinions on, and experiences with, technology are context-specific and dynamic.[6] Furthermore, understanding the contexts in which a technology is used, and its purposes, is key to the design of transparency and trustworthy mechanisms that do not violate social norms.

In the case of PETs, one way to improve transparency is to communicate to users the protection properties of the technology through explanations or descriptions. While warning messages have been thoroughly explored and evaluated in the academic literature,[7] [8] much less is known about how to

[4] Dodds, L. (2018), 'Who do we trust with personal data?'.
[5] Baker, A. (2023), 'Data assurance: Building trust in data'.
[6] Nissenbaum, H. (2004), 'Privacy as Contextual Integrity'.
[7] Akhawe, D. and Felt, A.P. (2013), 'Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness'.
[8] Felt, A.P et al. (2015), 'Improving SSL Warnings: Comprehension and Adherence'.

design effective explanations of what protection properties do or how they work. Taking encryption as an example, Abu-Salma et al.[9] argue that a high-level explanation of a secure communication tool as 'end-to-end encrypted' is too vague to inform lay-users of that tool's security properties. At the same time, Ruoti et al.[10] note that making the ciphertext visible to lay-users after the encryption takes place increases user trust in an encrypted communication tool. Furthermore, Alaqra et al.[11] found that structural explanations improved users' trust in encryption and their satisfaction, regardless of users' tech expertise, while functional explanations improved users' comprehension. This led Alaqra et al. to recommend combining both types of explanations. These findings hint at the highly contextual nature of the relationship between transparency and trust for more traditional types of PETs. In this research, we collected similar in-depth empirical evidence about emerging PETs[12] – namely TEEs – which had not been previously explored.

This dearth of evidence is partly due to the novelty of the adoption of some of these technologies, rather than the theory that underpins them. That said, it was important for us to first consider whether our previous working understandings of the concepts of both trust and transparency held in the case of PETs.

# What is trust in the context of PETs?

An organisation voluntarily adopting a technology (disregarding situations where it is forced to by regulation, and, likewise, disregarding profit incentives) will base its decision on a number of factors, including trust in the technology and the people and organisation providing it. For PETs, due to current levels of market adoption,[13] a primary consideration for prospective adopters at present remains whether the proposed solution solves a specific problem and can bring the organisation novel, otherwise unachievable, value. This research attempts to move beyond this initial obstacle to consider the factors affecting the next stage of PET adoption, namely how a specific product or vendor can demonstrate trustworthiness and the extent to which providing meaningful transparency to adopters helps this.

In social contexts, trust is defined as an individual's 'belief and expectation that all members in an exchange will act in a socially

[9] Abu-Salma, R. et al. (2017), 'Obstacles to the Adoption of Secure Communication Tools'.
[10] Ruoti, S. et al. (2016), '"We're on the Same Page": A Usability Study of Secure Email Using Pairs of Novice Users'.
[11] Alaqra, A.S. et al. (2023), 'Structural and functional explanations for informing lay and expert users: the case of functional encryption'.
[12] Centre for Data Ethics and Innovation (2021), 'PETs Adoption Guide [BETA]'.
[13] ODI (2022), 'Privacy Enhancing Technologies: Market Readiness, Enabling and Limiting Factors in the UK public sector'.

appropriate manner, and will not behave opportunistically by taking advantage of the situation'.[14] Mayer et al.[15] collate definitions of trust across literature, modelling trust decisions as the willingness of an individual to take a risk and work with others towards a goal, confident that others will help them to do so and will not, instead, exploit the individual.

In the context of PETs, where organisations are looking to adopt technologies to protect their clients' and their own sensitive data, these definitions can be extended. To integrate a PET provided by a third-party into its technological stack, the adopting organisation must be willing to make itself vulnerable to possible consequences that might be incurred through the use of the PET. This involves not only accepting risks in the functionality and efficacy[16] of the PET, but also trusting that the PET providers 'are acting, or will act, in an ethical and socially desirable manner'.[17] Trust in the PET is demonstrated in the organisation's willingness to adopt it when aware of these risks.

This trust can be informed and motivated or dissuaded by a number of factors.

# Transparency

When an individual or organisation makes an adoption decision, they base it on their interpretations of information that is available and accessible to them. With limited information, these interpretations are likely to lead to suboptimal decisions that can cause either the under-adoption of a technology or harm to the decision-makers.[18] Transparency measures can be implemented to mitigate these effects, working to increase the quantity, availability, and accessibility of information so that individuals and organisations can make well-evidenced adoption decisions. In doing so, transparency can also enable accountability – a noteworthy motivating factor for trust.[19]

---

[14] Zucker, (1986) 'Production of trust: Institutional sources of economic structure', as cited in Garrison, G., Rebman Jr., C. and Kim, S.H,. (2016), 'An Identification of Factors Motivating Individuals' Use of Cloud-Based Services'.
[15] Mayer et al. (1995), 'An Integrative Model of Organizational Trust'.
[16] For the purpose of this study 'efficacy' of a PET is taken to mean the capacity by which the technology can perform the anticipated function (in this case, increasing usability of data while maintaining privacy and security).
[17] Garrison, G., Rebman Jr., C.M and Kim, S.H. (2018), 'An Identification of Factors Motivating Individuals' Use of Cloud-Based Services'.
[18] Adapted from the concept of asymmetric information in the field of Economics, explored in Akerlof, G.A. (1970), 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism'.
[19] Fox, J. (2007), 'The uncertain relationship between transparency and accountability'.

Like trust, transparency is well defined from a social perspective. In practice, numerous technological transparency measures exist through which developers and providers of technologies have attempted to provide information about their products, datasets, code or models. Examples include documentation and open sourcing (both aiming to provide information on a technology's inner workings), real-time measurements and remote attestation (providing up-to-date information on the functionality, privacy and security of a technology), and formal verification (aiming to provide objectively measured guarantees of functionality, privacy and security).

In the field of Artificial Intelligence (AI), documentation-based transparency measures have been developed and introduced for both models[20] and their underlying training datasets.[21][22] While researchers aim to 'establish and promote […] foundations for transparency to pave the path for [the development of] systems and datasets that are responsible and benefit society',[23] working in such a fast-moving, high-visibility field gives them unique insight into types of transparency and how people interact with them. When seeking to understand focus group participants' perceptions of transparency efforts, Pushkarna, Zaldivar and Kjartansson[24] observed:

> *Despite the diverse backgrounds of participants across studies, the shared dominant perception was that transparency artifacts [sic] were ironically opaque. The opacity in documentation, quite simply, increases when language used is technical, dense, and presumptive of a reader's background, making it difficult for non-technical stakeholders to interpret.*
>
> **— Pushkarna, Zaldivar and Kjartansson[25]**

[20] Mitchell, M. et al. (2019), 'Model Cards for Model Reporting'.
[21] Data Nutrition Project (nd), 'The Dataset Nutrition Label'.
[22] Pushkarna, M., Zaldivar, A. and Kjartansson, O. (2022), 'Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI'.
[23] Pushkarna, M. and Zaldivar, A. (2022), 'The Data Cards Playbook: A Toolkit for Transparency in Dataset Documentation'.
[24] Ibid.
[25] Pushkarna, M., Zaldivar, A. and Kjartansson, O. (2022), 'Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI'.

Here, the researchers identify one of the inherent challenges associated with transparency efforts: they do not support decision-making if they are not made meaningful to decision-makers. A non-technical decision-maker is unlikely to have immediate use of technically-dense documentation that may be produced by basic transparency efforts.

In this regard, Haresamudram, Larsson and Heintz's three levels of transparency provide a review of the types of transparency that are necessary to (meaningfully) build trust.[26] While initially designed for AI, they can be extended to the wider context of general technology:

1. **Algorithmic transparency:** Underlying methods and decisions are broken down into ways that are understandable to humans. However, this does not necessarily mean they're easily interpretable, making algorithmic transparency most useful to technical experts and regulatory bodies.

2. **Interaction transparency:** More useful to laypersons/non-technical experts, interaction transparency informs on the primary effects of the adoption of a technology, including its benefits and costs as well as its requirements and constraints.

3. **Social transparency:** Similarly, social transparency provides wider information on a technology, including its potential positive and negative perceptions amongst different audiences, its governance and accountability measures, and its potential interactions with law and regulation.

The three levels noted above acknowledge that different types of transparency will serve different, but complementary, purposes to contribute towards more meaningful trust. Although the framework proposed by Haresamudram, Larsson and Heintz was initially conceived for the purpose of developing trust in AI, when it is broadened to general technology, it can serve as a solid foundation for similar considerations for PETs.

Furthermore, these types of transparency may take a variety of forms beyond typical means such as documentation. Technical transparency measures, like the reporting of real-time measurements, remote attestation and formal verification can provide algorithmic transparency by demonstrating what is going on under the hood of a PET. This can include information that decision-makers can use to inform their trust in the PET, and therefore their willingness to adopt it.

[26] Haresamudram, K., Larsson, S. and Heintz, F. (2023), 'Three Levels of AI Transparency'.

Through this report, we explore the extent to which transparency measures employed by PET providers motivate trust. We focused our research on a single type of PET: TEEs and their surrounding architectures. TEE providers could see adoption acutely impacted due to poor trust dynamics and therefore employ transparency measures to counteract them, thus making them a suitable PET to explore for this study.

# Trusted execution environments

TEEs are a type of hardware-based security that provides hardware isolation. This can be used by software processes that have strict sensitivity or privacy requirements so that the processes can operate on untrusted devices (eg, rented servers from cloud providers).[27] Technically speaking, TEEs operate using segregated parts of the device's computing power: for example, a server running with 16 CPU cores could have two of these cores segregated from the others when the TEE is booted. On this segregated compute, the TEE hosts a secure enclave kernel (like a virtual machine) that is built from a trusted binary image and runs an operating system that the user of the TEE can use to run their private and sensitive operations.

The rest of the untrusted device is blind to the processes occurring within the enclave, meaning no malicious actors with access to the device can observe or interfere with the operations that the TEE user wants to keep secret. TEE hardware is generally accompanied by a set of software instructions that are used to boot the TEE and the kernel safely; TEE hardware and software are hereafter both referred to as TEE infrastructures.

Modern TEE infrastructures have additional privacy and security features that enhance their hardware isolation capabilities. For example, some have systems to keep the CPU usage of the processes running in enclaves private from the resource monitoring of the untrusted devices they operate on, thereby making side-channel attacks (where attackers identify private processes by observing the resources being used in the device) impossible.[28] Likewise, sophisticated key exchange mechanisms guarantee the safety of information as it is transported between the user and the TEE, while authentication mechanisms provide assurances that the enclaves are booted correctly and uncompromised. The user's sensitive data, and the processes applied to it, is therefore stored with a significantly greater degree of privacy.

[27] Geppert, T., Deml, S., Sturzenegger, D. and Ebert, N. (2022), 'Trusted Execution Environments: Applications and Organizational Challenges'.
[28] Sasy, S., Gorbunov, S., Fletcher, C.W. (2017), 'ZeroTrace: Oblivious Memory Primitives from Intel SGX'.

As a result, TEEs enable 'confidential computing'.[29] Documented use cases for confidential computing include multi-party computation[30] and data localisation,[31] but the predominant, emerging use case for TEEs is in cloud computing.

When a client organisation is using servers in an external cloud service, it is running its operations on its data upon untrusted devices; the client does not own the servers (rather, the cloud providers do), and will have varying physical access or extensive monitoring capabilities to ensure that the devices are safe and uncompromised. Faced with such a prospect, clients operating with sensitive data that have strict privacy requirements may be apprehensive of the risk that this basic cloud computing brings. However, with an advanced TEE infrastructure, they can be assured that, even if their operations are happening on cloud servers millions of miles away, their data and their code is being kept segregated from the untrusted devices on an authenticated kernel.[32]

Of course, this architecture is not only business-to-business. Apple Intelligence has a sophisticated TEE architecture behind it (called Private Cloud Compute)[33] that aims to assuage concerns about how user data is used to train AI models. Since the mobile devices on which user data is collected do not have the computational power to train or fine-tune a user's personal Apple Intelligence model, the data must be sent to one of Apple's servers in a data centre for the computation to be performed. As the data includes text messages, emails and photos, users (and legislators) require the utmost assurance that it is being processed on secure devices. So, Apple employs a TEE infrastructure to build hardware isolation and assure users that data about them is not being exposed and misused, neither by external actors nor Apple itself.

## Trusting TEEs

By facilitating confidential computing, TEEs enable organisations to access the benefits of cloud computing even when dealing with extremely sensitive data. On the end-user side, products that integrate TEEs into their

[29] Scapicchio, M. and Kozinski, M. (2024), 'What is confidential computing?'.
[30] Law, A. et al. (2020), 'Secure Collaborative Training and Inference for XGBoost'.
[31] Schmidt, K. et al. (2022), 'Mitigating Sovereign Data Exchange Challenges: A Mapping to Apply Privacy- and Authenticity-Enhancing Technologies'.
[32] Boivie, R. (2022), 'Strengthening cloud security with confidential computing'.
[33] Apple (2024), 'Private Cloud Compute: A new frontier for AI privacy in the cloud'.

architectures are known to bring 'increased comfort'[34] as a result of their strengthened privacy capabilities. These properties suggest that the decision to adopt TEE architectures is simple.

However, to adopt TEEs, an organisation must trust them. That is to say, in accordance with the definitions of trust above, *an adopting organisation must have faith in the efficacy and functionality of TEE architectures as well as a belief that the providers of TEEs will not behave in a malicious way.*

### On efficacy and functionality

Over the last decade, the privacy guarantees of TEE architectures have often come into question, with the most high-profile controversies involving Intel's Software Guard Extensions (SGX).[35] SGX is the set of instructions that are used when booting up a TEE on an Intel CPU, performing the segregation and encryption of computational resources to enable hardware isolation. Since its introduction in 2015, SGX has become ubiquitous in TEE discussions, both because of its uses (in remote computing, secure web browsing, and digital rights management[36]) and its documented vulnerabilities. These include, and are not limited to, the 2017 'Cache Attacks on Intel SGX',[37] the 2018 'SGXlinger' side-channel attacks,[38] the 2019 'SgxPectre' speculative execution attacks,[39] and the 2020 'Plundervolt' fault-injection attacks.[40] [41] Such a history with the most well-known TEE architecture might serve to dissuade the adoption of TEEs in general.

### On the behaviour of the provider

In almost all cases, an organisation looking to adopt a TEE will not be a hardware manufacturer and therefore cannot build its own TEE architecture. As a result, the organisation must rely on one offered by a provider, such as Intel. This presents an issue: 'a main drawback of TEEs is the use of a hardware component that is fully controlled by the manufacturer',[42] meaning a TEE user relinquishes a certain level of control of the architecture to the provider. If the provider were to be malicious, it could exert this control to read the user's sensitive data or code and use it

---

[34] Musale, P. and Lee, A. (2023), 'Trust TEE?: Exploring the Impact of Trusted Execution Environments on Smart Home Privacy Norms'.
[35] Intel (nd), 'Intel® Software Guard Extensions'.
[36] Ibid.
[37] Götzfried, J. et al. (2017), 'Cache Attacks on Intel SGX'.
[38] He, W. et al. ( 2018), 'SGXlinger: A New Side-Channel Attack Vector Based on Interrupt Latency Against Enclave Execution'.
[39] Chen, G. et al. (2019), 'SgxPectre: Stealing Intel Secrets from SGX Enclaves Via Speculative Execution'.
[40] Murdock, K. et al. (2020), 'Plundervolt: Software-based Fault Injection Attacks against Intel SGX'.
[41] More can be found in Nilsson, A., Bideh, P.N. and Brorsson, J. (2020), 'A Survey of Published Attacks on Intel SGX', but more have been found since.
[42] Bouazzouni, M.A. et al. (2017), 'A Card-less TEE-based Solution for Trusted Access Control'.

for its own purposes. For an organisation to adopt a TEE architecture, it must have confidence that this situation will not happen.

This research aims to explore how the above considerations manifest in the minds of developers and security experts, and how transparency measures can be used to mitigate them. To do so, we focus on the trust paradigms surrounding a single TEE architecture with multiple transparency features: Google's Project Oak.

# Project Oak, by Google

Project Oak[43] is a software platform being developed by Google that allows users of TEE hardware to boot a TEE and launch an enclave on it in a transparent way, providing verifiable (or falsifiable)[44] claims about the TEE so that the user has as much information as possible regarding how their data and code is being treated by it. This transparency can, in turn, potentially address the trust issues with TEEs mentioned above for certain actors, which might encourage greater adoption and uptake of TEE infrastructures.

Oak is designed with three main transparency features, but at the time of this report's publication it remains under development, and this information is subject to change. At time of writing these features include:

## Open sourcing of code

As a starting point, Oak's entire code base is available for any interested parties to view, contribute to, and scrutinise on a verified GitHub repository.[45] Open sourcing shares similarities with efforts, such as Data Cards and the W3C Data Catalog Vocabulary,[46] in providing publicly available information (in this case, code and accompanying documentation) that enables interrogation and suggestions for amendment by adopters, security researchers, or interested third parties. This transparency aims to demonstrate that the provider of the technology is not doing anything – either accidentally or purposefully – that can be considered as compromising its users and their data.

[43] Google (2024), 'Project Oak'.
[44] Santoro, T. (2024), 'Falsifiability in confidential computing: A philosophical approach'.
[45] Google (2024), 'Project Oak'.
[46] Albertoni, R. et al. (2024), 'Data Catalog Vocabulary (DCAT) - Version 3'.

## Attestation reporting (remote attestation)

Attestation reporting is a function used to verify that both the TEE hardware and the enclave running on it have been booted correctly and are running uncompromised. Oak's architecture attests that 'the enclave application is running [on] up-to-date and correctly configured TEE [hardware]'[47] by cryptographically signing a hash of the TEE's binary, endorsing that the TEE is aligned with expectations. This evidence is supplemented by a set of endorsements from the manufacturer of the hardware that the TEE is running on, who sign a chain of certificates to state that the TEE has been set up on a legitimate CPU. Only the manufacturer can sign these certificates, therefore verifying that the enclave is running on their own hardware.

As a result, this architecture works on a split system, requiring the signed endorsement of both Oak and the manufacturer of the hardware (eg, Intel for Intel SGX/TDX TEEs, or AMD for AMD SEV-SNP). By doing so, the system distributes trust: the attestation reporting mechanism cannot be compromised unless both parties collude to do so.

Attestation reporting provides transparency on the status of both the device and kernel when both are in use, giving the user up-to-date information on the security of their data, the code and the device upon which it is all being run.

## Transparent release

The binaries used to boot enclave kernels on the TEE are the most likely sources of backdoors and other compromises, and are often authored or upkept by the open-source community. In the Oak infrastructure, any binary release or update is appended to an external, append-only log, alongside details of the author. This log is visible to the entire internet, and works to ensure that, if a compromised binary is released or updated, it and its author are both visible on a public log.

This assumes an 'ecosystem of verifiers' that monitors the public log and validates each new release. These verifiers are designated as 'key opinion formers' (KOFs). This is similar to a bug bounty-hunting service,[48] taking each release in the log and reproducing it to verify that it is uncompromised – or, if it is found to be compromised, reporting it quickly. Theoretically this could limit the ability of a malicious actor to insert a

[47] Google (2024), 'oak / README.md'.

[48] Bug bounties are a typical means by which companies can incentivise the reporting of identified vulnerabilities by individuals or organisations, as opposed to these being kept for later use or sold to prospective malicious actors.

backdoor into an enclave application, assuming that a KOF identifies and flags any vulnerability first.

Overall, transparent release provides transparency on the safety of kernels and – when they may not be safe – provides means by which to detect and report.

Oak combines these transparency features with a set of technical security mechanisms, including a sophisticated key exchange mechanism[49] that facilitates the secure exchange of data between the user and their kernel processes through the untrusted host device. Oak also allows users to keep the Trusted Computing Base (TCB)[50] small by using Oak's own Virtual Machine (VM) firmware and restricted operating system, in a bid to mitigate concerns from even the most privacy- and security-conscious potential adopters.

[49] Google (2022), 'oak / docs / images / BasicFlow.png'.
[50] Google (2024), 'oak / README.md'.

# This research

## Objectives

This research examined the relationship between trust and transparency in PETs, and the implications this has for their adoption. Particular focus was placed on TEEs and Oak,[51] the latter being a piece of TEE infrastructure designed to increase security and trustworthiness of TEEs by employing transparency measures. We explored the perception of these technologies from the perspectives of software developers and security researchers, two groups whom we identified as the primary influencers ie those whose opinions would inform the reception of Oak upon its release.

Oak offered us a chance to examine the extent to which transparency measures can translate to greater trust from developers[52] who might use, or be considering using, TEEs. This gave us the opportunity to consider the limitations of Oak's transparency measures towards addressing underlying general scepticism of TEEs and PETs.

We therefore structured our research based on the following overarching research questions:

1. Are there barriers to trust in Oak among developers and security researchers – and if so, what are they?
    - 1.1. What measures are necessary to incentivise developer adoption of Oak?
    - 1.2. What role does transparency, or lack thereof, play in developers' scepticism towards Oak?
2. How could Oak's transparency features affect developers' willingness to adopt the technology?
    - 2.1. Can similar features be employed across technological contexts to encourage adoption?

[51] Google (2024), 'Project Oak'.
[52] Throughout this research we have primarily engaged with 'developers', by which we mean individuals within organisations that are responsible for the design of data architectures that enable the secure use of sensitive data.

# Methodology

We began this piece by undertaking desk research on trust and transparency in the context of technology and PETs. During this phase, we also engaged with researchers working on TEE architectures to establish a comprehensive understanding of the trust dynamics surrounding TEEs. These researchers were selected for their dispassionate, critical and independent voices. This initial research informed the design and implementation of three data collection activities:

- A HotPETs workshop session at the 2024 PETs Symposium,[53] in which we presented an interactive slide deck with live polling of the audience (which was primarily comprised of academics) that focused on the motivating factors for trust and mistrust in PETs. The live poll was followed by open discussion with audience members.
- A two-stage in-person workshop with invited developers and security researchers.
    - In the first session, we used flashcards to prompt discussions regarding the motivating factors that influence participants' TEE trust and adoption decisions.
    - In the second session, the developer team for Oak presented their key transparency features, and participants discussed each in an open forum under the Chatham House Rule.
- Interviews with experts from within the PETs domain, in which we asked them to scrutinise the transparency features of Oak and provide their opinions of their effectiveness in motivating trust.

All views expressed in these data collection activities were personal and not representative of the participants' organisations.

Through these activities, we collected qualitative data, upon which we conducted thematic analysis to answer the research questions listed above. Throughout the course of our data collection activities, we engaged with more than 150 participants. The results of this thematic analysis can be found in the following section of the report.

For a full account of the data collection methods, including a detailed account of the development and delivery of the research workshop, please see the appendix on our data collection methodology.

[53] For details, see the HotPETs program

# Results

In this section we provide analysis of the findings gathered through the course of our research activities. These findings were grouped into four overarching themes that link to the research questions, focusing on trust in certain PETs – in this case, TEEs and Oak – and how this trust is motivated and dissuaded by a number of factors, including transparency. This analysis is a composite of the findings from all the research activities that we undertook: the interactive session at HotPETs, the workshop on motivating factors underpinning PETs adoption and features of Oak, and interviews with subject matter experts.

## 1. Transparency measures only motivate trust when they are meaningfully implemented

*Participants noted that the inclusion of some transparency measures like open sourcing are only the first steps towards a product being transparent enough to command their trust. They are not sufficient on their own.*

During the workshop, participants expressed their opinions on the trustworthiness that can be motivated by transparency measures taken by PET providers. They came to a general consensus that a fully transparent, open-sourced PET is not inherently more trustworthy to them than a closed-source counterpart because open-sourcing alone does not command trust. Rather, as an interviewee made clear, open-sourcing is a 'necessary but not sufficient condition' for their trust in a technology, explaining that open-sourcing is necessary because it allows the validation of the entire codebase, but insufficient because, by itself, open-source code is generally inaccessible to those without subject matter expertise.

While open-sourcing facilitates scrutiny of the code, interviewees expressed doubts that organisations looking to adopt a certain PET can be expected to have the time, human resources, and overall acumen to read the PET's entire codebase and make an informed decision regarding its safety and trustworthiness. Instead, workshop participants suggested that some organisations, likely small ones or those working in less data-sensitive industries, might rely on the 'many eyes' model of safety validation – the idea

that, because the PET's codebase is open-source, others can scrutinise it, so after a certain length of time, it will not have any vulnerabilities, either accidentally or purposefully, included in code. Other participants pointed out that this model does not provide 100% guarantees, noting the recent example of the 'XZ Utils backdoor' attack that affected Linux distributions like Debian.[54]

Rather than relying on the open-source community in this way, adopting organisations might prefer another organisation to carry out the validation of the codebase. Participants suggested these validating organisations should be strictly third-party, with no relation to the PET providers (avoiding potential bias in the validation) or the adopting organisations (meaning the adopting organisations hold no liability if the PET is found to be unsafe post-adoption). The HotPETs audience preferred this third-party verification to come from academic peer-review processes or government entities. On the other hand, workshop participants expressed a strong distaste for both, drawing from their experiences regarding the meaningfulness of academic reviews when it comes to industry and their personal views of government entities as technological decision-makers. Rather, workshop participants regarded verification of security best came from industry professionals they could trust who had already adopted the technology and had published their experiences in white papers.

Through this validation process, whatever methodology it might take, open sourcing encourages trust by reducing the asymmetry of information between provider and adopter. An interviewee remarked that '[the field of PETs] is very much a lemon market in that regard', with literature in the field of privacy economics, like Tsai et al.[55] and Acquisti et al.[56], finding that the adoption of PETs is often hindered by potential adopters being in positions of imperfect or asymmetric information. Transparency features like open sourcing make privacy information about PETs more salient and accessible, enabling trust and encouraging adoption.

But again, that trust requires the validation of the open-source technology; the PET being open-source is insufficient by itself. During the HotPETs session, an audience member made it clear that, in their experience, potential adopters' 'trust markers' (a person's requirements/criteria that must be filled for them to trust something) can be wildly different to those the providers expect them to have. In this example, a provider therefore cannot make its technology open-source and expect adopters to trust it. Instead, the provider should take meaningful steps to make it easier for the adopter to implement

[54] James, S. (2024),'FAQ on the xz-utils backdoor (CVE-2024-3094)'.
[55] Tsai, J.Y., et al. (2011), 'The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study'.
[56] Acquisti, A., Taylor, C. and Wagman, L. (2016), 'The Economics of Privacy'.

and verify the open-source code. This should be accompanied by clear and proper documentation, annotated and clearly structured code, and – at a minimum – a willingness to be contacted for information or demonstration of processes. Only then are the fruits of open-sourcing borne.

That said, it is worth noting that while a willingness to engage is a commendable step that is appreciated and should be encouraged, this can paradoxically invite criticism from some actors if they feel their request has been neglected. While this issue was raised in both the workshop and interviews, there was acknowledgement that a level of appreciation exists that not all requests can be addressed, or prioritised in a manner that will appease everyone. As such, the benefit of demonstrating a willingness to meaningfully engage was deemed a greater order of value to the community than the cost associated with unsettling certain individuals.

# 2. When it comes to trust in PETs, principles often trump efficacy

*Participants did not often base their trust evaluations and adoption decisions on objective assessments of a PET's efficacy – rather, their thinking was often guided by their own principles and subjective perceptions regarding the technology and, most of all, its provider.*

When workshop participants discussed the types of organisational providers of PETs they found easier to trust, they remarked upon the concessions they felt obligated to make in their experiences. One group presented an inherent distrust of 'big tech' companies, especially those with business models relying on targeted advertising, noting that such organisations present a threat 'too big to ignore' when being incorporated into the technological stack of an organisation working with sensitive data. Some participants however argued that 'big tech' companies should be trusted because they have 'more to lose' from either accidental or malicious controversy when it comes to privacy, given their size and social standing. Alongside the relative ease of clearing 'big tech' companies through due diligence and compliance checks (in contrast to the difficulties when working with relatively unknown organisations or the open-source community), participants felt that this meant the PETs these companies provide are adopted by even the most data-sensitive operations, despite any misgivings about their trustworthiness.

Importantly, assessments of trustworthiness were, as above, often based on the provider rather than the functionality of the PET in question. Participants believed PETs developed or hosted in non-Western countries are under-adopted simply because of this fact, where efficacy does not matter in the face of geopolitical risk.

There are, however, two sides to this coin. Workshop participants and interviewees alike pointed out that historical vulnerabilities can actually motivate trust in a PET, purely from an ideological perspective and despite efficacy concerns that the vulnerabilities could raise. Speaking about scenarios like Intel SGX, where a number of vulnerabilities have been found over its lifetime and Intel has taken measures to remedy them, participants said '[this history] shows that people are using the PET and finding vulnerabilities, and that the manufacturers receive them well with prompt patching, which builds trust in a way'. However, they noted that especially visible in the case of SGX are entrenched beliefs, where non-technical experts heard about a history of vulnerabilities and formed unwavering negative opinions of the technology, and refused to adopt it.

Entrenched beliefs[57] are common motivational factors for decision-making problems found in privacy literature.[58] They were particularly visible in HotPETs workshop participants' distaste towards state actors as technology providers or attestors. In this instance, they formed these opinions without consideration of the potential benefits of government entities as independent public-good providers, with participants instead making reference to and focusing on historical overreaches of intelligence agencies in the world of cybersecurity to justify their stance.

The HotPETs audience suggested other, non-efficacy-related reasons that could dissuade their trust in a PET. In agreement with workshop participants, multiple audience suggestions concerned the PET provider's business model and potential malicious incentives when working with sensitive data – in one suggestion, indicating that a provider that generates its income from donations or even a paid subscription service is much preferred to a provider that earns revenue from the use or sale of data for targeted advertising. Similarly, endorsements from, or affiliations with, divisive political figures and organisations that 'have violated privacy in the past' were suggested as a major factor that could dissuade adoption.

So, efficacy is only part of the story when it comes to adoption of a PET. Trust in the provider is generally based on the principles of the adopter – their concern with big-tech, their wariness of state actors, and their distaste

[57] Entrenched beliefs are often referred to as 'anchoring effects' and 'inertia' in behavioural economics.
[58] Redmiles, E.M., Mazurek, M.L. and Dickerson, J.P. (2018), 'Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions'.

towards historical vulnerabilities. However, whether these views are entrenched or not, it remains challenging to quantify the extent to which the effects of these principles can be mitigated by technical transparency measures.

# 3. Oak's transparency features have strengths and limitations

*Technical transparency features change the stakes of a trust decision. With good documentation, easy monitoring and visible metrics, efficacy and functionality concerns can be quelled, as adopters can see how the technology works, both before and after they adopt it. Alongside open-sourcing, attestation reporting, and transparent release, adopters can see more of what the provider of the PET is doing – introducing accountability and therefore motivating trust.*

*When discussing the specific features present in Oak, workshop participants and interviewees explained their viewpoints regarding the extent to which the above holds true.*

## Open sourcing

With regards to Oak's open-source nature, workshop participants and interviewees revisited their opinions expressed in the first theme: Oak's open-source nature does positively impact trust, but only if the open sourcing is made meaningful. They pointed out that Oak does actually accomplish this by providing extensive, visually informative documentation in project-specific, explanatory file formats (READMEs) across multiple folders and subfolders, but perhaps more could be done as the technology is developed to ensure that this documentation meets the trust markers that Oak's intended audience will have.

Interviewees also spoke about the extent to which Oak is open-source. Appreciating that the entire Oak codebase is publicly available (and stating that any technology that is only half open-source is as good as closed source), they discussed how the technology's full commitment to open-source, ie, allowing contributions and pull requests from the open-source community, inspired greater trust in Oak by demonstrating a level of respect towards the open-source ecosystem. They noted that there should be guarantees in place stating that Oak will continue to have this commitment to open source in the future.

## Attestation reporting

From the perspective of an individual concerned about the security of Oak, the attestation reporting mechanism theoretically works to quell concerns about where code or data is being sent and whether components of the stack are compromised. This should therefore encourage trust by verifying the efficacy of Oak as a PET.

In practice, interviewees pointed out that the certification approach used by Oak might not motivate trust from sceptics. Although the attestation reporting feature was described as a 'nice to have' by one interviewee, they noted that it limits the technology from operating in a 'zero-trust architecture',[59] in that, 'at this point, you have to trust someone'; any trust in Oak derived from the attestation report certificates necessitates trust in the certificate signees. Another interviewee defined this further, stating that they would need to trust the security of the attestors' signing keys and the overall measurement process – but most importantly, they would have to trust that the signees, Oak, and the manufacturer of the hardware are not signing certificates maliciously or otherwise incorrectly.

Furthermore, although the attestation reporting mechanism in Oak resembles 'distributed trust' models by having multiple parties, it does not function in the same way. In this case, the two signee parties are not 'checking each other's homework', ie, verifying that their counterparts are signing their statements correctly. Consequently, as one interviewee noted, malicious or incorrect signatures will not necessarily be visible and reported to organisations that use the technology. This limitation in the attestation reporting architecture means that it might not motivate trust from the most sceptical, privacy-aware decision-makers, nor organisations with more stringent regulatory or normative practices, such as public sector organisations.

However, the HotPETs audience, workshop participants and interviewees all wholeheartedly believed that the introduction of third-party attestors could allow the attestation reporting mechanism of Oak to motivate greater adoption. 'It depends [on what kind of third party],' one interviewee stated, 'and who [a potential adopter] would like to trust'. Aside from government organisations, the interviewee suggested attestors could take the shape of independent not-for-profits. Alternatively, the role could be assigned to competitors of the PET providers. While conflicts of interest would have to be negotiated, the interviewee reflected that one already exists in the current attestation reporting architecture (Oak's signature that the kernel is correct), and on balance, any sort of third party would be preferable to that.

[59] Kindervag, J. (2010), 'No More Chewy Centers: Introducing The Zero Trust Model Of Information Security'.

## Transparent release and key opinion formers

Oak features third-party attestation in the form of transparent release.

The involvement of Key Opinion Formers (KOPs) had upsides and downsides for workshop participants. The participants appreciated that this protocol worked to distribute trust: between the attestation reporting and transparent release, there is attributable blame for anything that goes wrong in the infrastructure. Therefore if there were a privacy breach or other compromise, at least one of the attestors (the manufacturer, Oak, KOPs) should discover it, and if they do not, they can be held liable for any damages that occur.

However, participants also noted that in their experience, the KOPs being from the open-source community is a limitation. This is simply because open-source work often comes from diverse communities of individuals, sometimes anonymous and other times from unknown countries or organisations. Therefore, while in some instances it may be beneficial to receive diverse input, which may signal a willingness on behalf of an organisation to broad scrutiny, these properties could make compliance checks difficult for leading industry organisations that might want to adopt Oak. Simply put, this may serve to undermine this attempt at transparency as they may not be able to trust the provenance of all verification statements made by KOPs.

# 4. Technical transparency measures only address certain concerns for certain actors

*Transparency features like those in Oak may provide visibility and accountability, but only in specific, technical ways that might appeal to only certain audiences.*

Adoption requires trust in both the PET's efficacy and its provider's behaviour. Technical transparency features like those in Oak enable the former by allowing adopters to see what the PET is doing, attesting to its security and safety while providing means for third parties to provide verification. However, these features are less likely to address trust in the provider, which is typically contextual and perception-based.

An interviewee came to a similar conclusion to workshop participants, stating that the motivating factors for a person's trust in something are entirely based on who they are and what context they are in.

For example, when the decision-maker in charge of adoption is not a technical expert, or is otherwise unfamiliar with the principles the PET is intended to address, trust in the provider is all the more important. Workshop participants pointed out that compliance officers and lawyers 'don't expect to understand the principles behind the technology', concluding that the effectiveness of technical transparency features might be limited when it comes to encouraging adoption if they are not tailored to a diverse range of decision-makers.

Similarly, the HotPETs audience suggested many non-technical attributes of a PET provider that might dissuade an adopter's trust, such as political endorsements, historical affiliations with unreliable organisations or intelligence agencies, and holding a monopoly. These factors cannot be wholly addressed by technical transparency features, and will affect the trust people have in a PET and its adoption.

# Discussion

## Making it easier to say ‘I trust you’

The themes above indicate that there are obstacles to trust in Oak which, although mitigated to varying extents by technical transparency features, are based on subjective perceptions and principles just as much as objective assessments of the efficacy and credibility of providers.

This will, unfortunately, likely serve to limit the adoption of Oak. Organisations holding the most sensitive data and using the most sensitive operations will base their decision to adopt Oak on their trust in the technology, and if it falls short of their decision-makers’ subjective requirements for that trust, the technology’s efficacy and transparency features hold limited power: as a consequence, the organisation will be unlikely to adopt Oak.

As found in the workshop and at HotPETs, these subjective requirements are wide ranging; some potential adopters can find a plethora of ways to not trust, and not adopt, a technology. As an example, one HotPETs participant suggested they would not trust anything with ‘flashy marketing’. For certain individuals who hold entrenched, sceptical attitudes on a technology or its provider, very little can challenge their worldview; it will always be difficult for them to say ‘I trust you’.

This might be the result of their ‘neuroticism’ or ‘chronic privacy attitudes’, two factors that have been found to affect people’s perceptions of their vulnerability and, as a result, their intention to adopt self-protective behaviour,[60] [61] which in this case takes the form of aversion and scepticism toward trusting a PET provided by a third party. Sceptical individuals may also over-value the risks (or under-value the benefits) of adoption, perhaps influencing others to do the same. PET providers are restricted in their ability to counteract these motivations and have their technologies be widely adopted.

Oak’s transparency measures demonstrably help in this regard. As seen in theme three above, participants across our data collection activities found the combination of open sourcing, attestation reporting and transparent

[60] Brough, A.R. and Martin, K.D. (2020), ‘Critical roles of knowledge and motivation in privacy research’.
[61] Chennamaneni, A. and Gupta, B. (2022), ‘The privacy protection behaviours of the mobile app users: exploring the role of neuroticism and protection motivation theory’.

release measures to be a meaningful solution to trust problems in TEE architectures. While they pointed out their concerns in the precise technical implementations, they made sure to note that, as per Voltaire, 'perfect is the enemy of good'.[62]

In their current iteration, though, the transparency measures might not fully capture the trust of the most untrusting people, who have more difficult chronic privacy attitudes, neuroticism, and scepticism. To reach these people, effort must be made to ensure that each transparency measure is as meaningful as is feasible, as described in theme one above, by taking steps to ensure that they meet the trust markers of their audience.

Furthermore, as per theme four, they must meet the trust markers of all types of decision-makers, which might be a resource-intensive task. Sceptics are not always subject matter experts, especially in the world of privacy and security: Brough et al.[63] find that knowledge of, and literacy in, privacy matters has no impact on a person's willingness or refusal to adopt a privacy technology. In other words, a lawyer, for example, is no less likely to be a sceptic of a TEE architecture than a cybersecurity expert. Transparency features must therefore be meaningful across that spectrum of knowledge and literacy, being useful for, and inspiring trust from, as many people as possible.

However, the findings of theme two persist. Despite being presented with useful, meaningful transparency measures, a sceptic might still refuse trust and adoption on the grounds of their own, subjective, principles. The solution here, in all likelihood, is not technical.

One HotPETS audience member – a founder of a global-scale internet privacy technology – shared their solution to the problem of scepticism. In their case, their highly privacy-aware user base had very low risk appetites for their own data, especially regarding internet browsing. Although the founder's PET was built to perform exactly their required purpose – obscuring internet browsing data – potential adopters did not trust the PET provider not to spy on the data. This belief was not based on any objective evidence, but rather on scepticism arising from the chronic privacy attitudes and neuroticism of the user base. The founder told us:

[62] Voltaire (1772), 'Contes en vers (Voltaire)/ La Bégueule'.
[63] Brough, A.R. and Martin, K.D. (2020), 'Critical roles of knowledge and motivation in privacy research'.

> *[The potential adopters] were never going to be satisfied. They would keep asking questions, finding ways to not have faith in us [the developers]. Eventually, we just had to start saying 'don't worry about it, we know it isn't perfect, but you have to trust us'. We only made this statement effective by building that trust in our team, with community outreach, interaction, and co-design.*
>
> **— a HotPETs audience member, who was the founder of a global internet privacy technology**

# Conclusion

In this research, we set out to answer a set of research questions regarding trust in Google's Project Oak, and how its transparency features and other factors can motivate or dissuade it. While the research was therefore focused on Oak and TEE infrastructures overall, our findings can apply to all types of PETs and offer insight for technology more generally.

Over the course of our study, we used literature to define trust in the context of PETs, identify how transparency can be used to motivate it, and explore these dynamics with regard to TEEs and Oak. We then used qualitative data collected from three activities, and analysed it to understand the underlying themes and motivations behind discussions regarding trust and transparency in PETs.

Our findings provide empirical evidence on the considerations made by privacy-aware developers and researchers when they choose to trust, or not trust, a PET. This is important information that technology providers and academics can use to understand barriers to adoption of PETs and technology more generally. We find that distrust can be based on subjective perceptions and principles rather than objective analysis, and demonstrate the varying extents to which this is true across a number of socio-technical factors.

We also illustrate the dynamics by which transparency helps, or does not help, contribute towards trust in a technology. While Oak's transparency features have strengths, in general transparency must be made meaningful – for all types of decision-makers – to truly affect trust.

We discuss how holistic measures must be taken to enhance trust and therefore persuade adoption. These measures can include transparency features, but must also contain socio-technical methods to help align trust markers between PET providers and their target audiences.

## Recommendations

In summary, technological transparency is a necessary component for building trust, but is likely insufficient. Where possible and reasonable, providers should seek to maintain and advance the levels of technical transparency that they can provide to adopters, aiming to meet expectations where possible. These technical transparency measures must also be accompanied by wider considerations, such as their meaningfulness to all

kinds of decision-makers. This transparency should be complemented by community outreach and participatory co-design to ensure that it meets the trust markers of adopters.

We therefore recommend that further work is required to explore what the co-design of transparency might look like, and the exact mechanisms by which it can be used to enhance trust. As we proposed in our analysis, one area that should be considered is how attestation reports are scrutinised and verified. The different communities we engaged with through this research had differing opinions on the types of organisational actor they preferred to perform attestation, which suggests that additional co-design must have participation from different types of individuals and organisations. There are a variety of means by which this can possibly be conducted, whether by specific individuals or organisations, or by industry or open-source communities. As we have identified through our research, the word of specific actors will resonate more strongly with some than others, which will contribute towards the development of trust in the technology. Identifying the most suitable actor to undertake third-party verification or auditing is not, however, a straightforward and intuitive process and will depend on a number of factors.

Relatedly, as the PETs ecosystem continues to evolve towards a greater state of maturity, we believe that it will be important for developers to work closely with adopters – potentially by sector – to build mutually trustworthy technology. In encouraging this collaboration further upstream, the opportunity to develop the foundations for trust between vendors and adopters will be increased.

A final recommendation we propose is to undertake similar research on the transparency measures necessary for adjacent PETs to TEEs. Because PETs are highly context and purpose-specific – each requiring individual consideration – we have found in previous work that there are limits to which you can generalise across these technologies. For example, PETs have varying risk and threat models. As such, this piece of research would benefit from comparison with a similar analysis of another type of PET, which would both serve to identify commonalities between PETs and increase our ability to generalise or compare our findings.

Trust and transparency are topics that are brought up frequently in discussions on the adoption of digital technologies, across contexts and geographies. While this report is specific to the context of PETs, and focused on TEEs and Google's Project Oak, we hope that it provides meaningfully contributive evidence for the analyses of interactions between humans and technology, whatever that technology may be.

# Acknowledgements

This report would not have been possible without the generous support from the team at Google Research.

We are also incredibly grateful for the time and input provided by the many experts who contributed to the research activities that were undertaken throughout the course of this research. This includes the interviewees, workshop participants, and those who provided input, who include, but are not limited to:

- Irina Bejan, Openmined
- Dave Buckley, OpenMined
- Prem Eruvbetine, Google
- Sarah de Haas, Google
- Andreas Haggman
- Daniel Hugenroth, University of Cambridge
- Ben Laurie, Google
- Ben Moore, Department for Science, Innovation and Technology
- Tiziano Santoro, Google
- John Smith, Department for Science, Innovation and Technology
- Lacey Strahm, OpenMined
- Fredrik Strömberg, Mullvad VPN AB
- Naaman Tamuz, Bitfount

We would also like to express our thanks to our colleagues at the ODI who

have contributed their time and assistance to the creation of this written report.

# Appendices

## Data collection methodology

This section contains a fuller account of the activities undertaken as part of the data collection methodology. Our research sought to build an empirical evidence base of attitudes towards transparency measures that could increase trust in TEEs. The rationale for the selection of our four data collection methods was primarily driven by the need to gather qualitative data, derived from the experiences of different groups and their motivations towards the adoption of certain technologies. The richness of this qualitative data was required, as factors that motivate trust in something or someone are experiential and thus the result of a multitude of many interconnected factors, which require exploration through means that allow for participants to elaborate on the reasoning behind their motivations.

**These activities included:**

### Desk research

This review included consideration of TEEs and the concepts of trust and transparency, as they relate to the adoption of PETs. We focused on attitudes towards the adoption of these technologies by those who are likely to incorporate them into their technological stacks, exploring how their decisions are influenced by trust in the operational and privacy guarantees that accompany these technologies.

#### Workshop 1: HotPETs workshop session at PETs Symposium 2024 (n = +100 in person, with online attendees – all of whom paid a registration fee to attend)

In our session, we sought to engage with stakeholders from within the PETs community through a combination of a presentation of the findings from our desk research and a survey, followed by discussion with the audience.

The survey questions posed to the audience were based on our desk research. In these questions, we aimed to collect data from those working within the field of PETs – specifically on motivating factors that would influence their decision to trust and adopt a PET from a specific provider.

**Results of the survey questions are as follows:**

## To what extent do you agree with the following statements?

76 responses

*"What happens on my phone, stays on my phone"*


*"I can trust the privacy protection measures of typical consumer products to keep me safe"*


*"I consciously take measures to protect myself and my data often"*


## Do you think a developer would adopt a PET even if...

58 responses

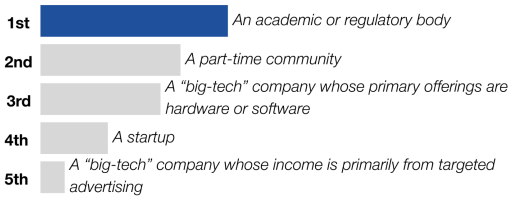
*... the PET slows down their product?*


*... the PET is monetarily expensive?*


*... it's hard to explain what the PET does, to end users?*

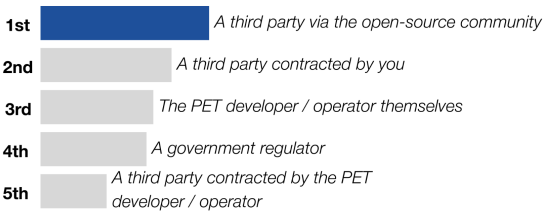

## What type of organisation do you trust most as developers and maintainers of PETs? Rank these options.

57 responses




## In an attestation protocol, whose signed statements would strengthen your trust the most? Rank these options.

46 responses




## If a PET was developed by a single organisation with no competition or substitute products, to what extent would your adoption decision be affected?

40 responses

### Workshop 2: Workshop on motivating factors underpinning PETs adoption and features of Oak (n = 16)

In parallel to planning for the HotPETs workshop, we designed a workshop to engage specifically with developers and security researchers who were familiar with PETs. This workshop was designed to complement the first workshop by inviting a smaller group of selected experts with specific expertise on TEEs, as opposed to the first workshop, in which we had no influence over the participants. The design of this workshop followed the structure of the research questions included in the section on the project objectives.

We split the half-day workshop into two sessions. The first was a group-based discussion that was prompted by cue cards that we had designed to spur discussion around a specific variable that related to motivating factors that might influence an individual or organisation's decision to adopt a PET.

The second session consisted of a presentation of the transparency features of Oak by members of the team at Google Research, followed by opportunities for the participants to provide their feedback and reflections on the features.

Participants were recruited from within the existing network of contacts from within the ODI, through recommendations from those within the ODI network and through efforts to identify participants through searching online. We sought primarily to recruit developers and security researchers with specific expertise in PETs and – where possible – TEE architectures. This narrowed the field quite substantially, given the specificity of these requirements. We also sought where possible for attendees to be based in, or near to, the UK, to increase the likelihood of in-person participation. All but one participant was able to attend in person on the day of the workshop. Nine participants were from industry, two were from academia, two from government departments, one from civil society and one from a media organisation.

## Expert interviews

Following the second workshop, we conducted a series of interviews to gain additional insights into the perception of the transparency features of Oak, as we had not managed to collect as much data that addressed research questions 2 and 2.1.

Again, our recruitment for these interviews was focused on speaking primarily with developers and security researchers.

# Biases and limitations

We are conscious of several biases and limitations that have a bearing for our research and eventual findings.

One noticeable limitation was the lack of diversity of research participants throughout the course of our activities. As noted in our data collection methods, above, we sought to invite developers and security researchers with familiarity of PETs and – where possible – TEE architectures specifically. Of the 16 confirmed participants, 14 were male. This is a much higher proportion of male participants than we aimed for. However, due to a combination of unavailability and lack of responses, we found ourselves with this heavily unbalanced ratio. It is worth noting that gender distribution in related sectors, such as cybersecurity, is widely acknowledged as disproportionate – it is estimated that women constitute only 11% of the cybersecurity workforce in the UK.[64]

Another limitation related to participants is that through this research, we explicitly sought to engage with developers and security researchers, rather than include prospective end users of Oak. We made this conscious decision given the current stage of adoption. As Oak is still in development and has not yet had extensive applications, we deemed it necessary to first focus our efforts towards gathering empirical evidence from those who would be most likely to interact with Oak and thus serve as early opinion formers. We did not engage with end users of Oak, but we believe that future research should focus on engaging with this group, once it is feasible.

A final limitation to acknowledge is the inconsistent response rate to the survey questions we presented to the attendees of the HotPETs workshop at PETs Symposium 2024. Appreciating that there would be attendees with varying levels of familiarity with TEE architectures, we decided not to make the answering of each question mandatory. As a result, we ended up with varying levels of responses to the questions posed to the audience, which limits our ability to draw comprehensive conclusions from the data gathered. Nonetheless, the incomplete answer sets still provided insights into the expectations amongst the expert community present at the PETs Symposium, which was useful in gauging the sentiment within the wider community.

A bias that we sought to acknowledge through the activities that we conducted was to account for the privacy attitudes of participants who took part in each of the activities. For this, we posed a simple question to

[64] Kerwick, V. (2024), 'The Underrepresentation of Women in Cybersecurity Leadership in the UK'

participants at the beginning of each of our activities. At the first workshop, this consisted of the statement: ‘I consciously take measures to protect myself and my data often’. From 76 responses, the modal responses were four and five (out of five, where five = strongly agree), meaning a significant portion of the audience believed it was highly privacy-motivated. As a result, we attempted to account for this in our analysis, acknowledging that those with whom we interacted were naturally inclined towards considering privacy matters comprehensively.

Prior to the second workshop, we requested that participants complete a pre-workshop survey featuring questions about their privacy attitudes. This included a question in which we asked to what extent they agreed with the statement: ‘I consciously take measures to protect myself and my data often’, where one = strongly disagree and five = strongly agree. Responses were somewhat low, with only six of the 17 participants completing surveys. Nonetheless, two respondents selected two, one respondent, three, two respondents, four, and one respondent, five. Overall, this suggests that of the responses received, there was a slight tendency towards greater privacy awareness.

Finally, at the start of each of our expert interviews, we asked our interviewees a similar, simple question: ‘On a scale of one to 10, how often do you take privacy-preserving measures when it comes to you or your household?’. The lowest response that we received from our interviewees to this question was eight.

The intention behind asking these questions was primarily to acknowledge that we were engaging with respondents who were already reasonably privacy-conscious.