

medConfidential comments on the TREEHOOSE project report

The approach DARE UK's Trusted Research Environment and Enclave for Hosting Open Original Science Exploration (TREEHOOSE)¹ takes to safety is akin to opening a box of hand grenades and asking who wants to play catch, then claiming the project is "safe" because *you* decided to leave the pins in...

Key points:

- **Builds on an assumption of "data lakes" for *all* researchers²**
- **Explicitly designed to allow an analysis to be kept secret from both TRE admins and data controllers / custodians³**
- **Uses some *terms* from Data Saves Lives, while undermining their *policy intents***

"Data Lakes" as a start assumption (page 2, section 2.2.1⁴) rely upon a large dataset being available to all researchers, to be used without visibility and without mechanisms to support the integrity of the output checking process.⁵

The ability to run "encrypted" code (page 5, section 2.2) means no-one using this system can be – or claim to be – consensual, safe, or transparent. This functionality cannot satisfy the requirements of safe people, cannot satisfy Safe Projects, and cannot satisfy Safe Outputs on what will be – and what TREEHOOSE appears to *require* to be – Unsafe Data.

In simple terms, the output checkers will not be able to assess any such outputs for risk, as the secret code can change the outputs to mislead the output checker – who will have no way to know how any output was created, and whether any of the output is what it claims to be.

Putting AWS tools together in such a way as to break the Five Safes model so completely will have taken serious, considered expertise – or a decision to employ a form of policy ignorance that seems calculated to *appear* to satisfy Data Saves Lives and the Goldacre Review, without understanding why the goal of actually increasing patient confidence is important..

NHS bodies review data agreements and infrastructures as required to ensure all parties are acting legally, as well as with integrity. Assessments of the infrastructure can be made by (accountable) public bodies which are being asked to put data onto untrustworthy infrastructures.

Every researcher having their own private 'data lake' of NHS patient data is incompatible with trustworthy data use. Indeed, the patient data lake model that underpins TREEHOOSE – HDR UK's view being incompatible with consensual, safe, and transparent data governance by the NHS – appears chosen to engineer these safeguards out of the system.

Data auditors should, as part of their audit and due diligence functions, understand where that feature request came from, where it has been used, and what else a culture that believes it acceptable has infected throughout the system as designed – it is rare that cover-ups only affect one place.

¹ <https://dareuk.org.uk/sprint-exemplar-project-treehoose/> – noting the support of UKRI, HDR UK and ADR UK

² Because data minimisation is only for people who follow the law?

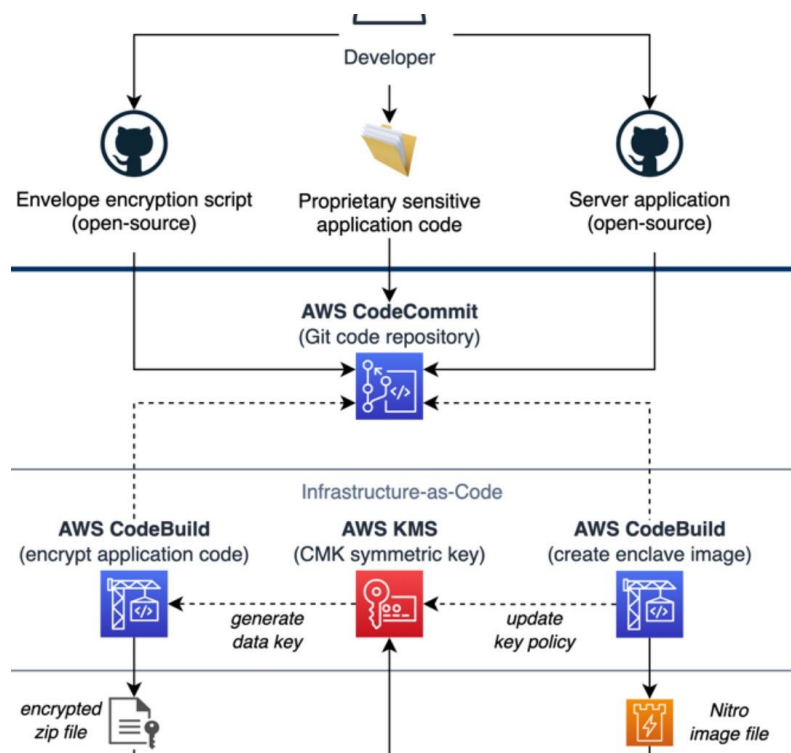
³ Which creates the basis for *unsafe* projects, *unsafe* outputs, and *unsafe* settings – returning to the unverifiable and untenable reliance on "safe" people and "safe" projects that has repeatedly failed in a dissemination culture.

⁴ Unless otherwise stated, all pages / sections referenced are in TREEHOOSE's end of project report, published here: <https://zenodo.org/record/7085505> (...until zenodo goes away, adds a barrier to public access, etc.)

⁵ Hence unsafe outputs.

Annex: TREEHOOSE evidence

Extract of Figure 2, 'Schematic of Secure Enclave capability within the TREEHOOSE architecture' (page 6, final report) noting "(encrypt application code)":



And the motivation for doing this was covered in a Dundee presentation:⁶

Secure enclave: resolving two conflicting requirements



We want to retain full control of our data



Commercial partner wants to maintain confidentiality of their analysis code

⁶ Link to presentation slides here: <https://discovery.dundee.ac.uk/en/activities/treehoose-trusted-research-environment-and-enclave-hosting-open-o>