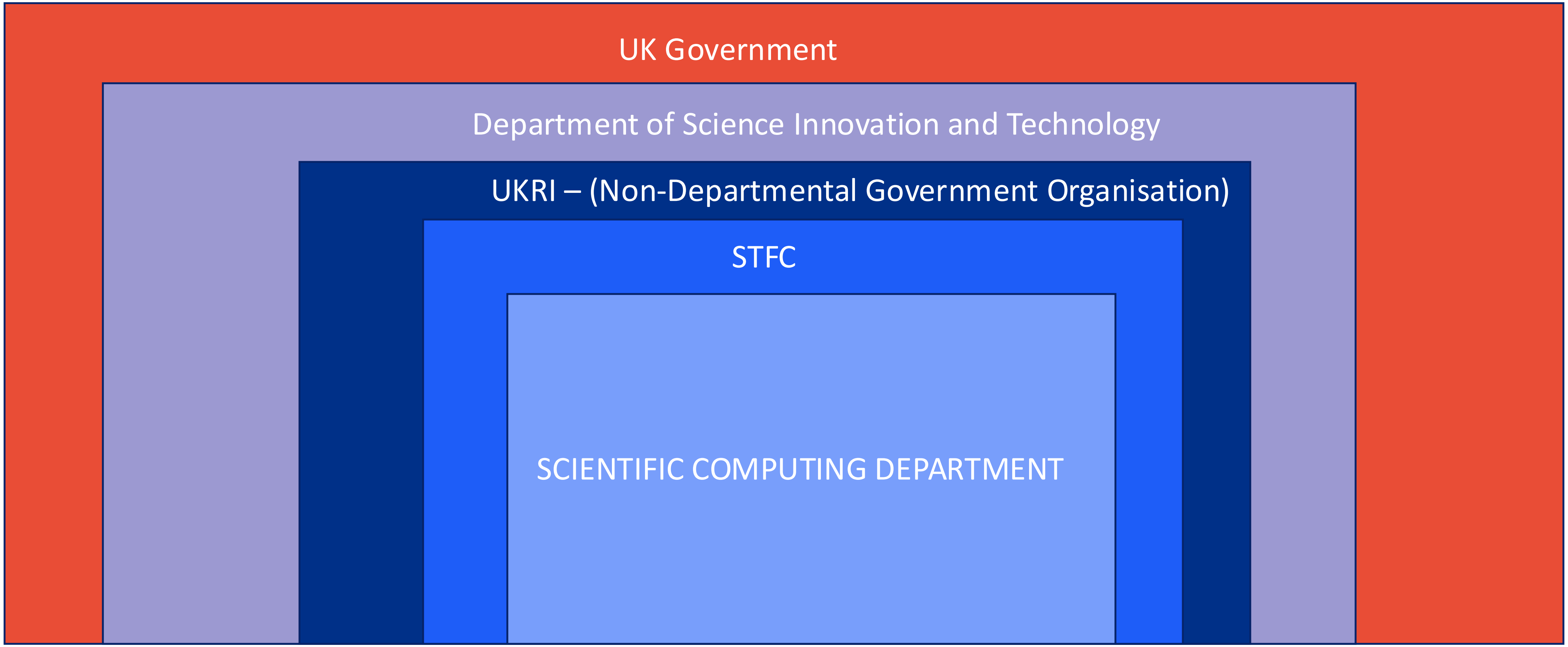
UKRI Science and Technology Facilities Council

GenAI Risk Management

Image © STFC Alan Ford

SCD Background



UK Government

Department of Science Innovation and Technology

UKRI – (Non-Departmental Government Organisation)

STFC

SCIENTIFIC COMPUTING DEPARTMENT

Background

- No GenAI risk work done being made available by
  - STFC
  - UKRI
  - DSIT
  - Government
- GenAI providers offer heavily biased risk analysis
- UKRI states that only web-based Enterprise CoPilot should be used
- GenAI in Project Management SIG created and use cases collected
  - These people have paid Enterprise CoPilot licences
  - Org mostly not project managers

What are we not doing

- Telling people “no”
- Assessing how useful GenAI is in reality
- Enforcing UKRI Policy
    - You have to use the web portal version of enterprise CoPilot
    - UKRI will not pay of enterprise CoPilot
    - Dept doesn’t have the money to pay for people to have enterprise CoPilot
    - 30 Project Managers across UKRI have a paid version
    - UKRI as a whole is 7736 people
    - The paid version is $30 a month
    - $232,080 per month for all of UKRI (£171,980 or €197,523)
    - UKRI themselves use an assortment of tools

What do we need to know

- What are staff *actually* doing?
  - Models
  - Work being done
  - Hosting location
- Has anyone else started addressing the exact same problem
- What are the risks Users and Management are concerned about

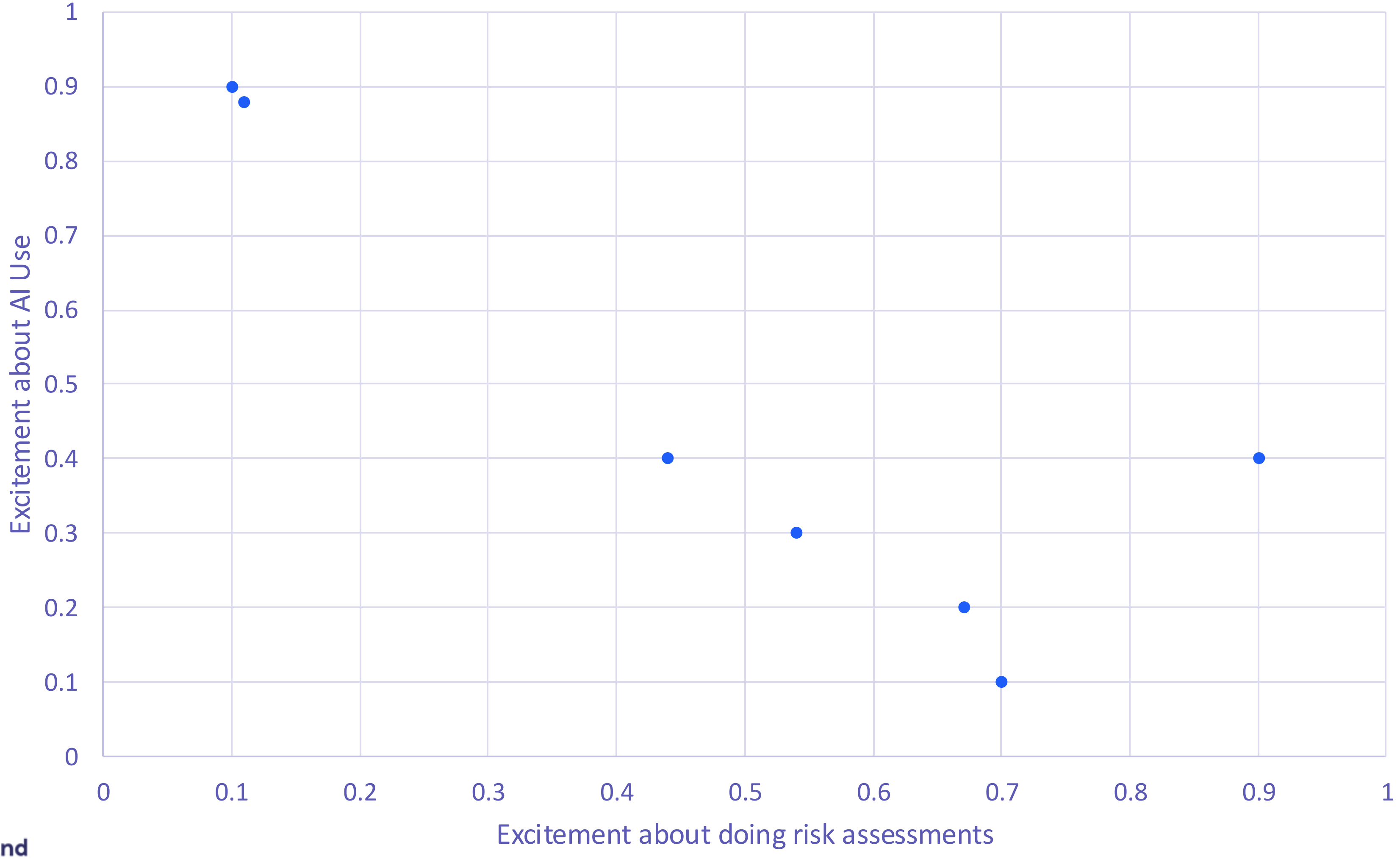
What are we worried about

- STFC IP infringement
  - Our IP being leaked
  - Pulling in other peoples IP
- Data Protection
  - Personal Data
  - Sensitive Data
    - OFFICIAL, OFFICIAL – SENSITIVE, SECRET and TOP SECRET
  - Allowed for Official and Official-Sensitive to be shared with Enterprise CoPilot Only
  - Secret and Top Secret are obviously not allowed to be shared with any LLM

Range of work being done

- Tier-1
- Data and Analytics Facility for National Infrastructure https://www.dafni.ac.uk/
- STFC Infrastructure
- Researchers
  - Computational Biology
  - Computational Chemistry
  - Computational Physics
- Project Managers
- Admin Staff
- Databases team

Feeling from talking to people



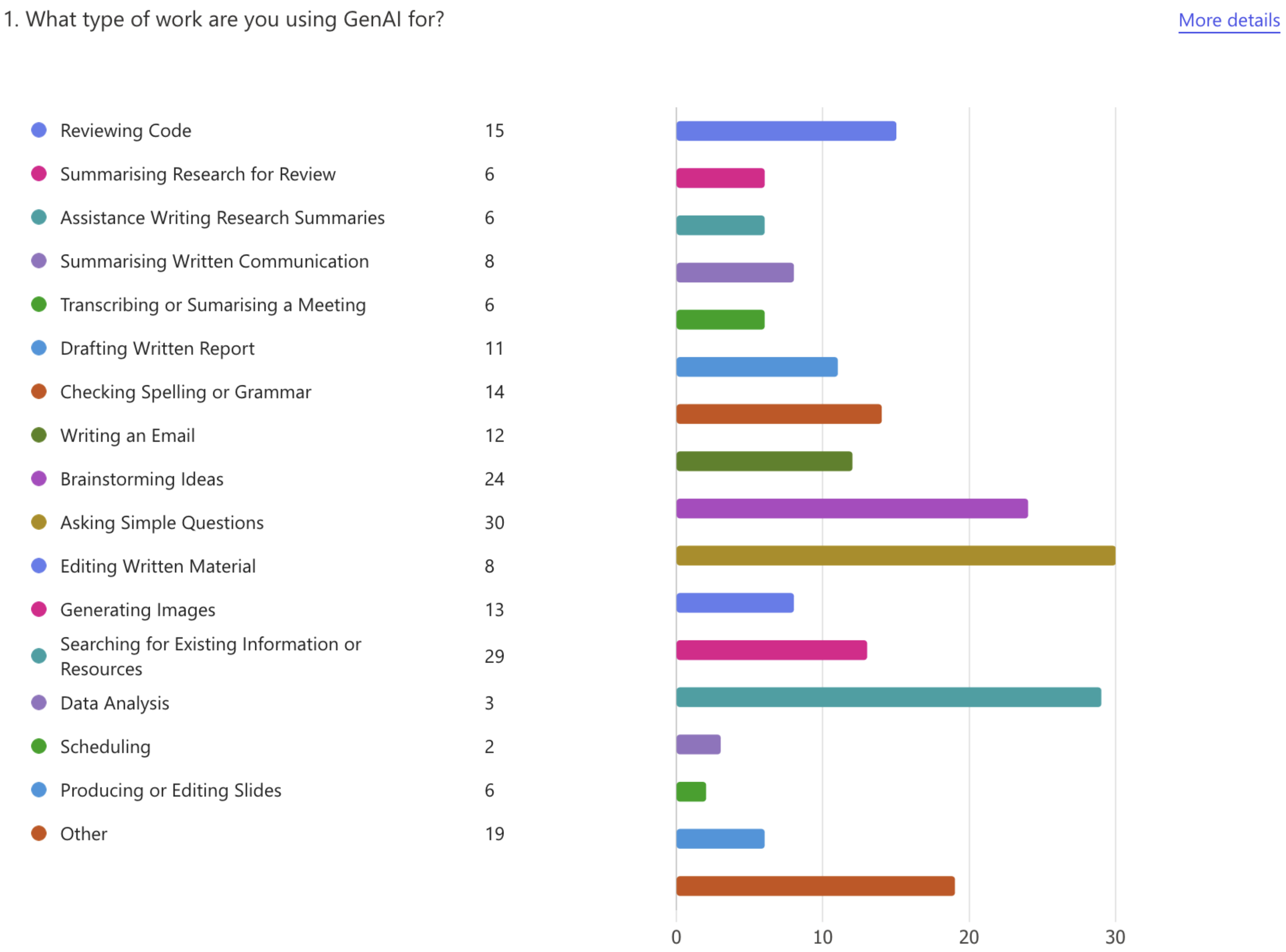
Science and Technology Facilities Council

Source – I made it up

Survey

- Fairly good response rate
  - 1 email with form link sent last week
  - 65 replies from 311 members in the department about 20%
- Only 7 questions
  - What are you using GenAI for?
  - How do you access your chosen model?
  - Is any used data classed as personal sensitive under GDPR?
  - What is the highest data classification of used data?
  - Do your inputs/questions contain any STFC IP?
  - Do you directly import any AI generated content into project outputs?
  - Any other comments?
- Anonymous Responses

Survey



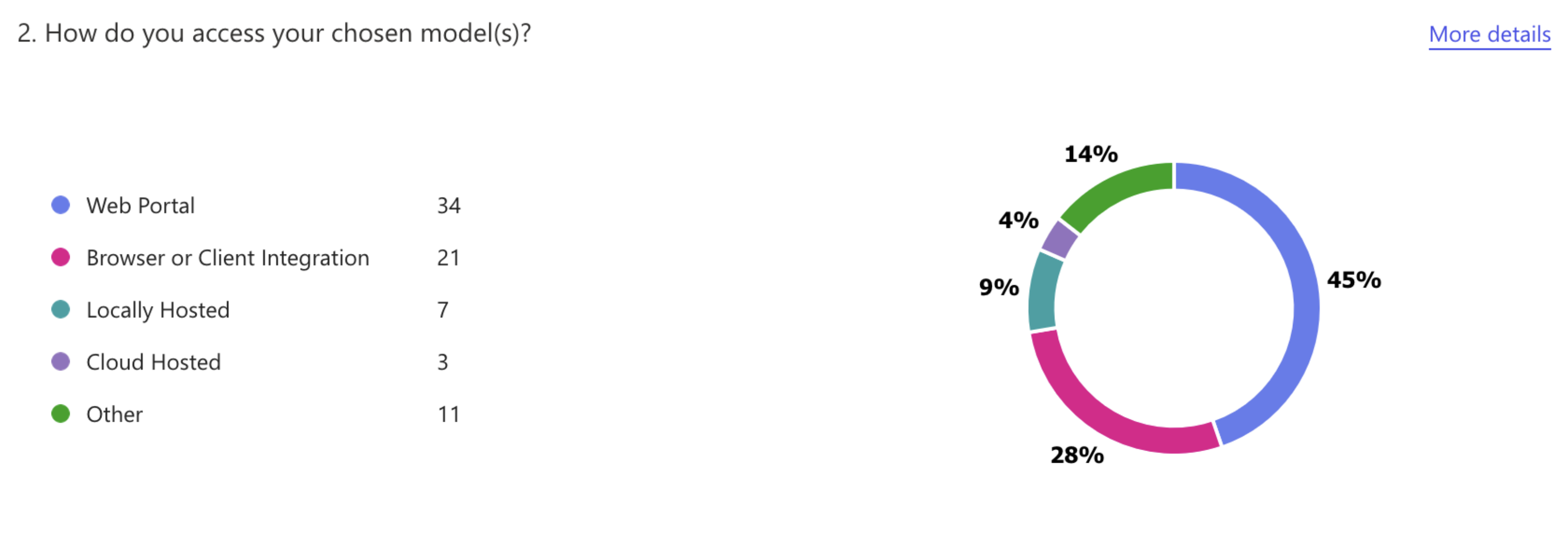
1. What type of work are you using GenAI for?

●	Reviewing Code	15
●	Summarising Research for Review	6
●	Assistance Writing Research Summaries	6
●	Summarising Written Communication	8
●	Transcribing or Sumarising a Meeting	6
●	Drafting Written Report	11
●	Checking Spelling or Grammar	14
●	Writing an Email	12
●	Brainstorming Ideas	24
●	Asking Simple Questions	30
●	Editing Written Material	8
●	Generating Images	13
●	Searching for Existing Information or Resources	29
●	Data Analysis	3
●	Scheduling	2
●	Producing or Editing Slides	6
●	Other	19

Other responses mostly “I don’t use AI” and a few for writing scripts

Science and Technology Facilities Council

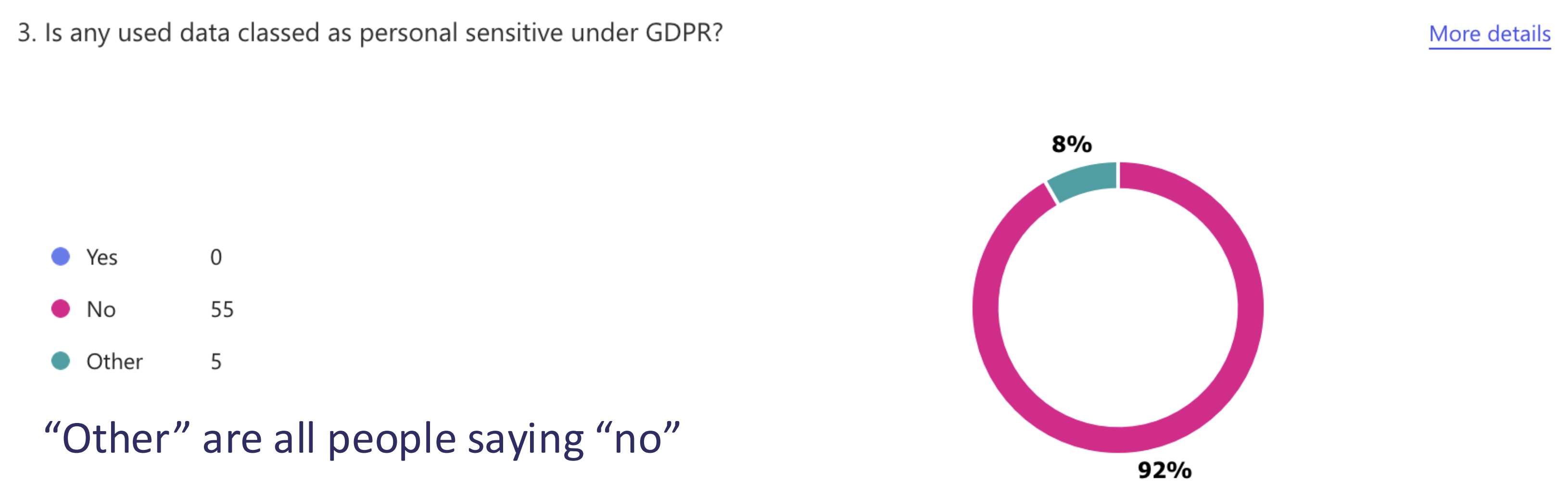
Survey



2. How do you access your chosen model(s)?

More details

●	Web Portal	34
●	Browser or Client Integration	21
●	Locally Hosted	7
●	Cloud Hosted	3
●	Other	11



3. Is any used data classed as personal sensitive under GDPR?

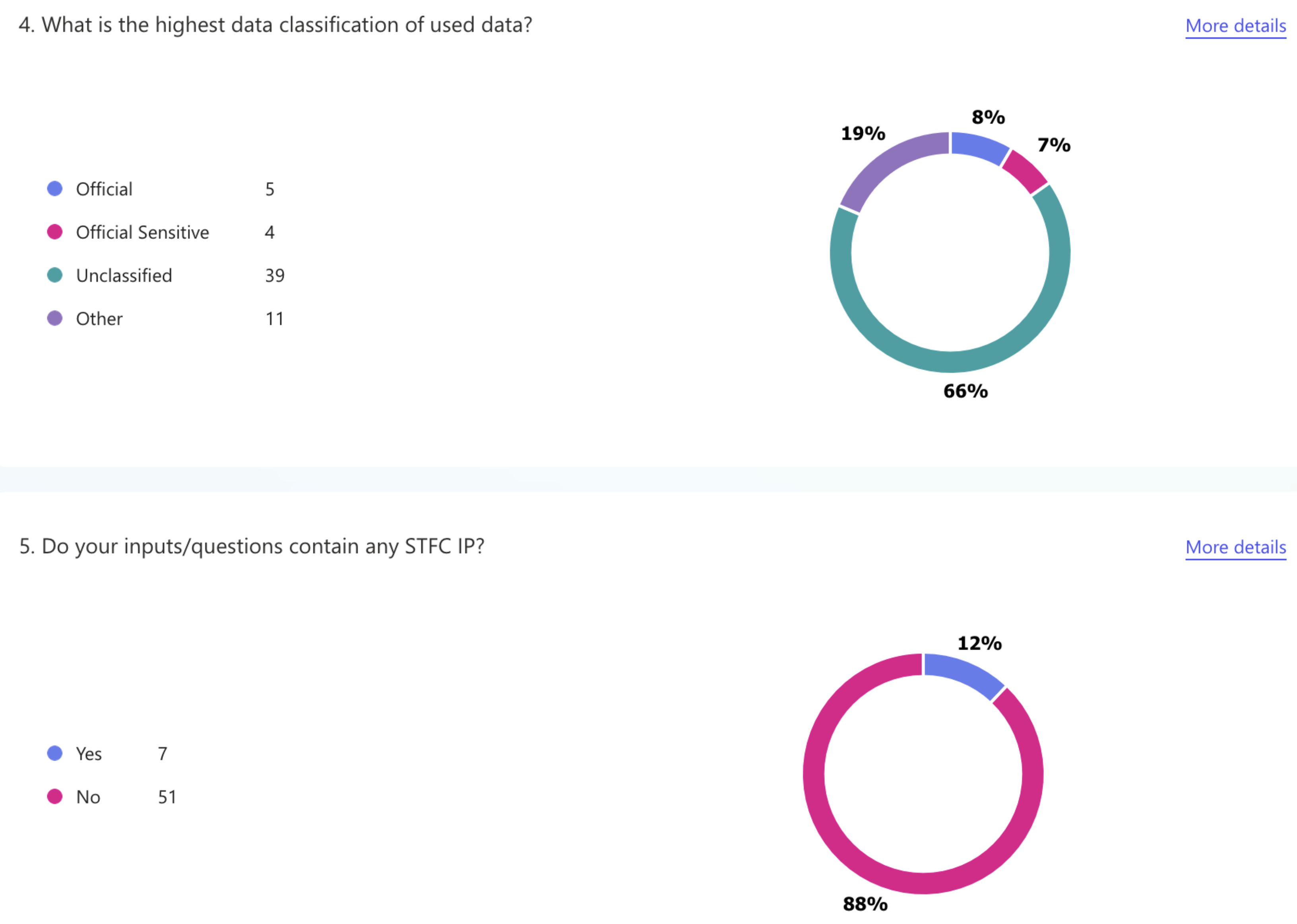
More details

●	Yes	0
●	No	55
●	Other	5

"Other" are all people saying "no"

Science and Technology Facilities Council

Survey



4. What is the highest data classification of used data?

More details

● Official	5
● Official Sensitive	4
● Unclassified	39
● Other	11

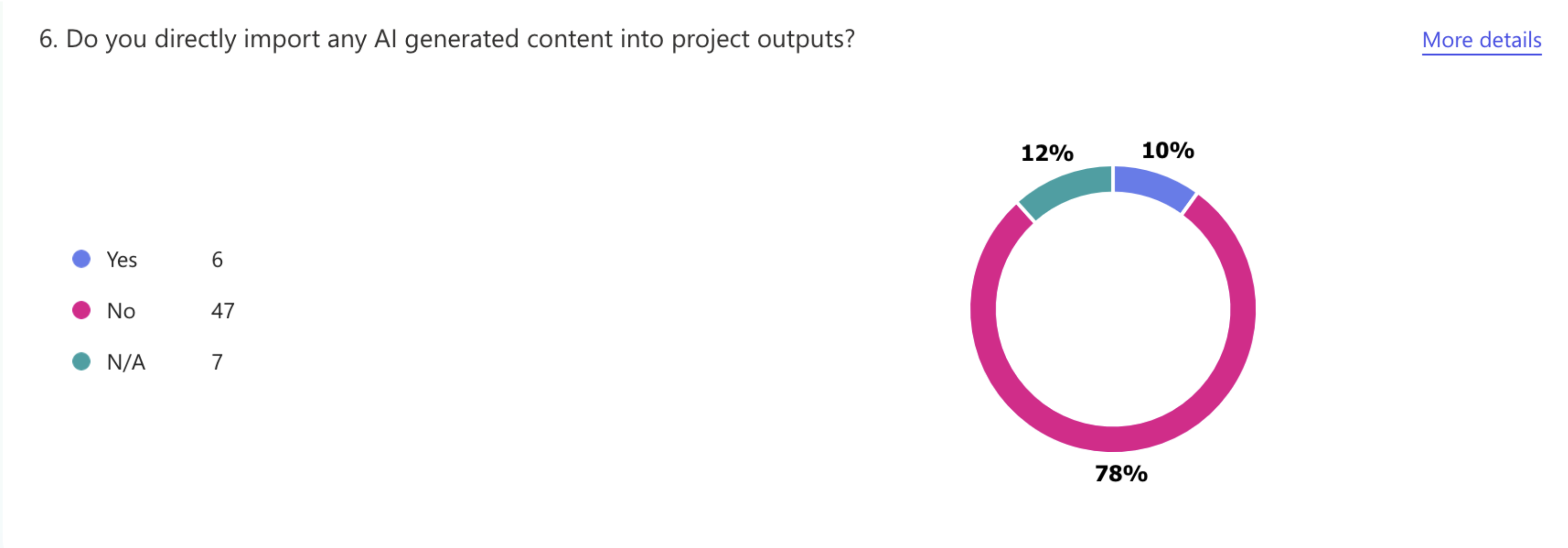
5. Do your inputs/questions contain any STFC IP?

More details

● Yes	7
● No	51

Science and Technology Facilities Council

Survey



6. Do you directly import any AI generated content into project outputs?

More details

●	Yes	6
●	No	47
●	N/A	7

- Q7 Range of responses
  - “I try to avoid the use of generative AI”
  - “I think it's getting to the point where not using GenAI is impossible; even the likes of spelling and granmmar checkers often incorporate it now.”
- Almost completely useless independently but may be useful if I go through individual responses
- I also had replies to my email from people sharing what they are working on or sharing opinions

Risk Assessments

- Already done a few
  - Github CoPilot, actually a mix of
    - OpenAI
    - CoPilot
    - Claude
  - On-Prem LLM, deployed with docker container onto Openstack VM
    - https://arxiv.org/abs/2402.19173
    - “StarCoder2-3B”
  - Early risk assessment version
    - Lots of input from project managers
    - New risk assessment shorter and more focused

The plan

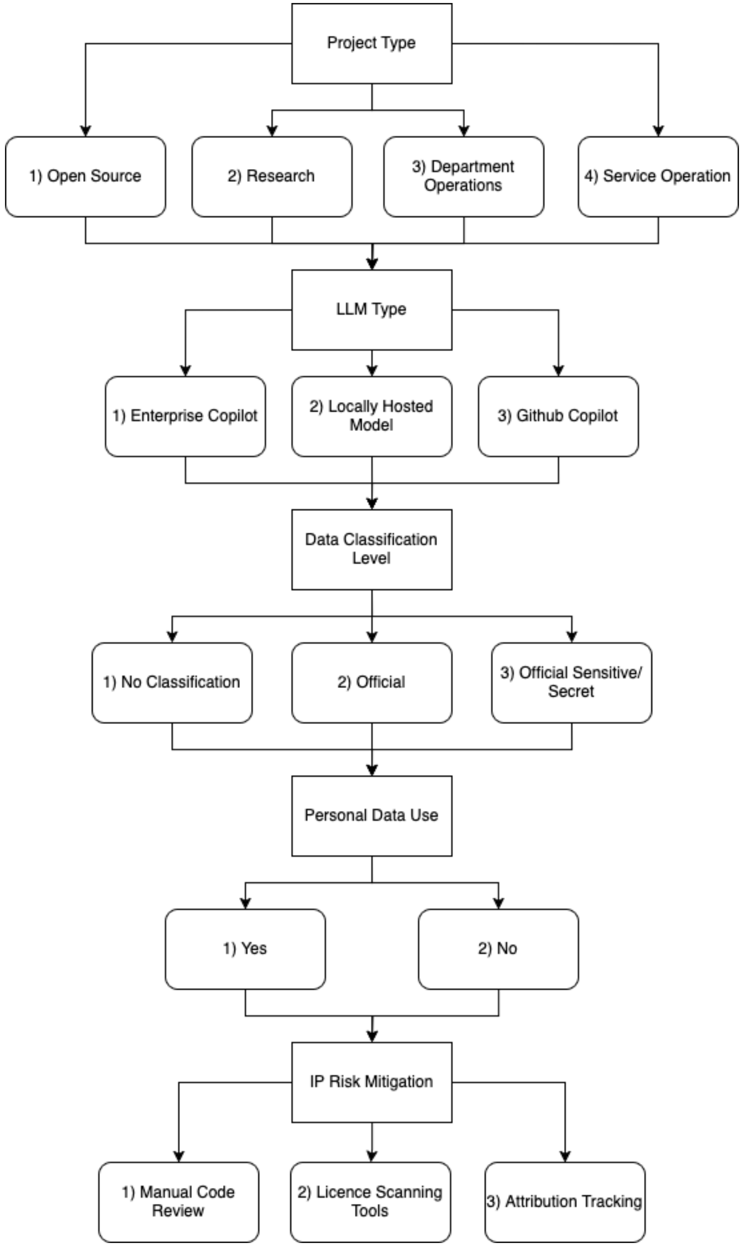
- Risk assessed by least senior person capable of owning risk
    - At STFC legal responsibility only possible from band F and upwards
    - Usually group leaders
        - Recent reorganization
        - Promotions frozen half way through
    - Risk assessments approved by “theme leaders” and department head
        - Themes are the next split under division
            - Range of sizes
            - Largest 45 people
            - Smallest is the security theme 3.5 permanent staff plus 2 graduates

How is it going

- Starting push once I have a gap (next Monday)
    - Time to support the effort for 3 weeks
    - Hopefully no services break
- Group leaders have been informed in the plan
- Specific meetings with AI Theme Lead
- Risk assessment Short so hopefully painless
    - Example risk assessments created to help guide

Long term plan

- I don’t want to do this longer than I have to
  - Asked to do the work by the Dept head, now we have a new Dept head
  - Group leaders create
  - Theme leads and department head approve
  - Security officer advises
- Part of department operations
- Hope that once most use cases have been done, we can move away from risk assessing everything
  - Diagram on the right was made to explain the hoped for end process
  - Can we just ask 5/6 questions and if someone has already risk assessed the same use case you just follow their risk assessment?



Science and Technology Facilities Council