Questions on notice from the Parliamentary Joint Committee on Law Enforcement:
Inquiry into law enforcement capabilities in relation to child exploitation
Public Hearing – 10 August 2023

1 PROOF HANSARD, pp. 5-6

Senator SHOEBRIDGE: Has Google undertaken any work to put in, for example, watermarks into product that's computer generated using Google tools or any other kind of indelible markings in the content that's produced to enable it to be traced? Is that part of Google's technological response?

Ms Cashman Kirstein: We have committed to doing that, yes.

Senator SHOEBRIDGE: But when you say 'committed to doing that', it hasn't been done yet?

Ms Cashman Kirstein: I don't know that I have the exact time line, but I believe that was part of the set of commitments that were announced at the White House a few weeks ago, along with other generative AI involved companies.

Senator SHOEBRIDGE: Could you, on notice, advise us where the implementation of that is up to and when it's expected to be implemented, including the other elements that were announced on that date?

Ms Cashman Kirstein: Absolutely.

Response: Google was proud to jointly commit with other AI companies to advance responsible practices in the development of artificial intelligence. These commitments represent a milestone in bringing the industry together to ensure that AI helps everyone, and they include a commitment to develop robust technical mechanisms to ensure that users know when content is AI-generated, which can help enable creativity with AI while also reducing the dangers of fraud and deception. These commitments will support efforts by the G7, the OECD, and national governments to maximize AI's benefits and minimize its risks.

While generative AI makes it easier than ever to create new content, it also presents new challenges for content trustworthiness. Two important approaches that we are investing in to address these challenges are watermarking and metadata.

Strengthening Content Trustworthiness through Watermarking. Watermarking allows for information to be embedded directly into content, even when an image undergoes some modifications. It is one tool in our overall toolkit to help indicate when content is created with Google's generative AI tools. Moving forward, we are building our generative AI models to include watermarking capabilities and other techniques from the start. Like other technical mitigations, watermarking has inherent limitations and could be manipulated or circumvented by adversarial actors. Still, watermarking is an important part of our approach to improve the integrity of the overall ecosystem.

Providing More Information Through Metadata. Metadata allows creators to associate additional context with original files, giving users who encounter an image more information. In the coming months, through our "About this Image" tool, Google Search users will be able to see important information such as when and where similar images may have first appeared, where else the image has been seen online, including on news, fact-checking and social media sites—providing users with helpful context to determine whether what they are seeing is reliable.

As we begin to roll out generative image capabilities, we will ensure that every one of our AI-generated images has metadata to provide users who come across them outside our platforms with valuable context. Creators and publishers will also be able to add similar metadata, enabling users to see a label in images in Google Search marking them as AI-generated.

---

2 **PROOF HANSARD, p. 6**

Senator SHOEBRIDGE: I just have one other question. Are you aware of the industry codes that are overseen by the eSafety Commissioner in Australia?

Ms Longcroft: Yes we are.

Senator SHOEBRIDGE: And you'd be aware that the eSafety Commissioner rejected the code for search engines because it didn't appropriately address the technological developments in generative AI—are you aware that that was the reason why it was rejected?

Ms Longcroft: Yes. In fact, we've been working very closely with the eSafety Commissioner on the online safety codes. It was an unprecedented coming together of industry to work with the eight codes and the eight different sectors of the tech industry. Those codes impact 35 of Google's various parts and, as I said, we've worked collaboratively with the eSafety Commissioner. All codes were accepted except for two, which the eSafety Commissioner explained, and the one that she sent back for further consideration around search tools did require some consideration to be given to generative AI.

We've worked across industry closely to propose provisions that would address the eSafety Commissioner's concerns with regard to generative AI and we are hopeful that the commissioner will find those acceptable. We will certainly continue to work with the commissioner and her team to ensure that any concerns in the field of generative AI are addressed.

Senator SHOEBRIDGE: What are the changes that are being proposed to that rejected draft?

Ms Longcroft: I don't have the particular details to hand, and that's currently before the eSafety Commissioner for her response, but I'd be happy to follow up with you on notice.

Senator SHOEBRIDGE: Alright, thank you.

Response: To correct the record, the search engine services code was not rejected. The eSafety Commissioner has reserved her decision on this code because the original draft (submitted in March) did not sufficiently capture recently proposed changes to search engines to incorporate generative artificial intelligence features. A revised draft search engine services code has now been submitted to the

Commissioner and includes a number of modifications including an updated definition and scope of “search engine services” that makes clear that any generative AI functionality incorporated within a search engine is captured by the commitments of the code.