Safeguarding Agency Data and Customer Information in the World of AI

September 10, 2025 – St. Paul, MN

Session Overview

- Introductions
- Crowdsourcing Exercise
- Responsible Use of AI
- Digital Forensics and AI
- Main Takeaways
- Q & A

Let’s meet our speakers:

Carl Medley II, Director of Security and IT Operations, eWorld

Zachary Rasmussen, Child Support Director, CGI

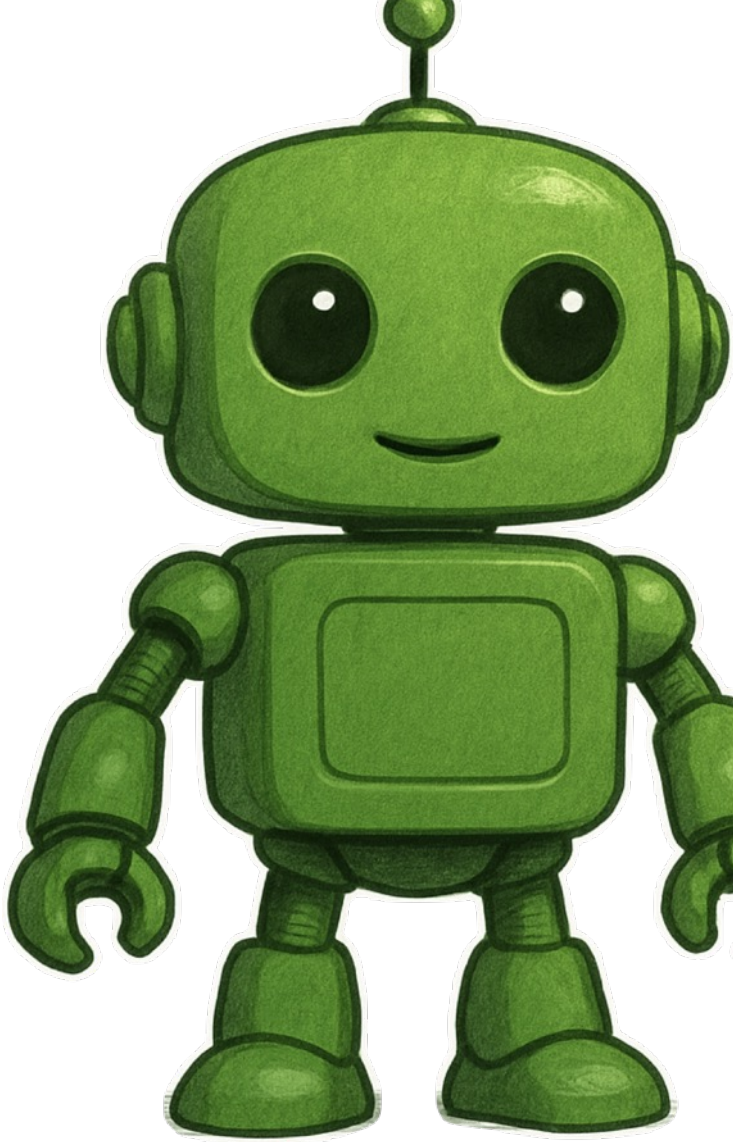
John Carney, Esq., Chief Technology Officer, Carney Forensics

Crowdsourcing Exercises

Morning Mood Assessment

Crowdsourcing Exercises

Technology
Loyalty
Assessment

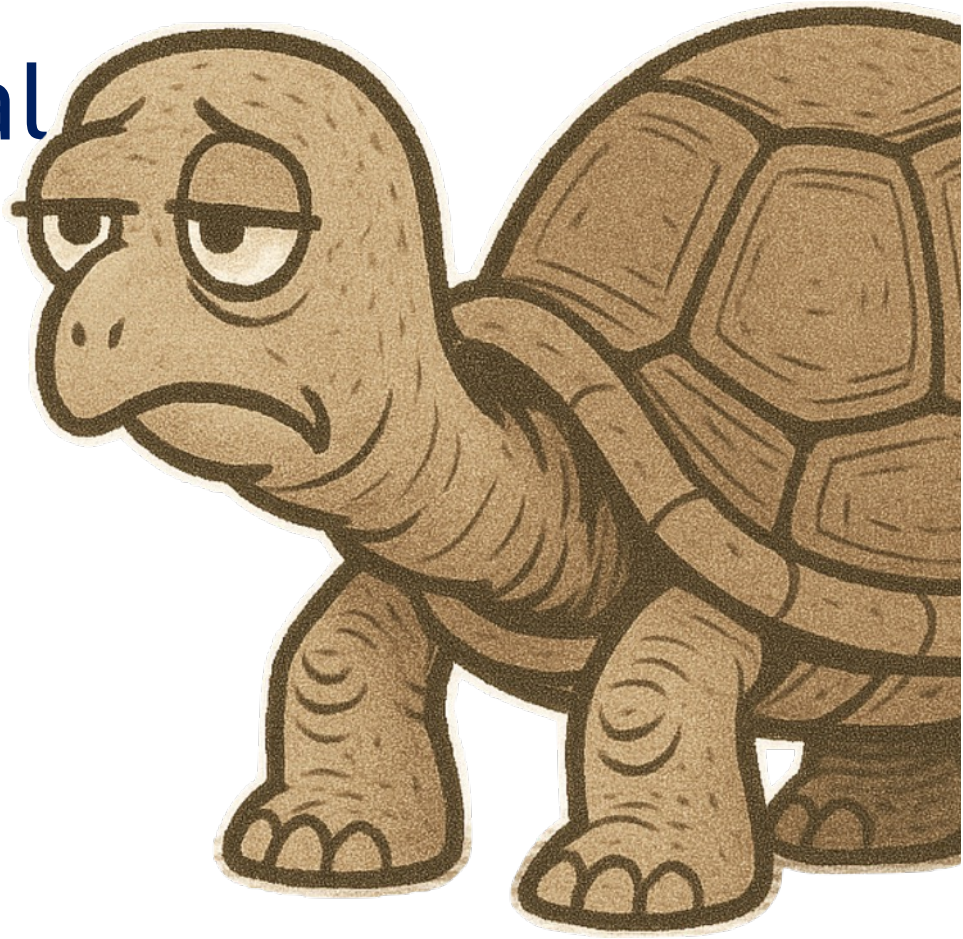


Crowdsourcing Exercises

Organizational Preparedness Assessment

Crowdsourcing Exercises

Organizational Response Efficiency Assessment



Crowdsourcing Exercises

Individual Information Security Awareness & Acuity Assessment

Responsible Use of AI

The Age of AI is here

From automation to creation

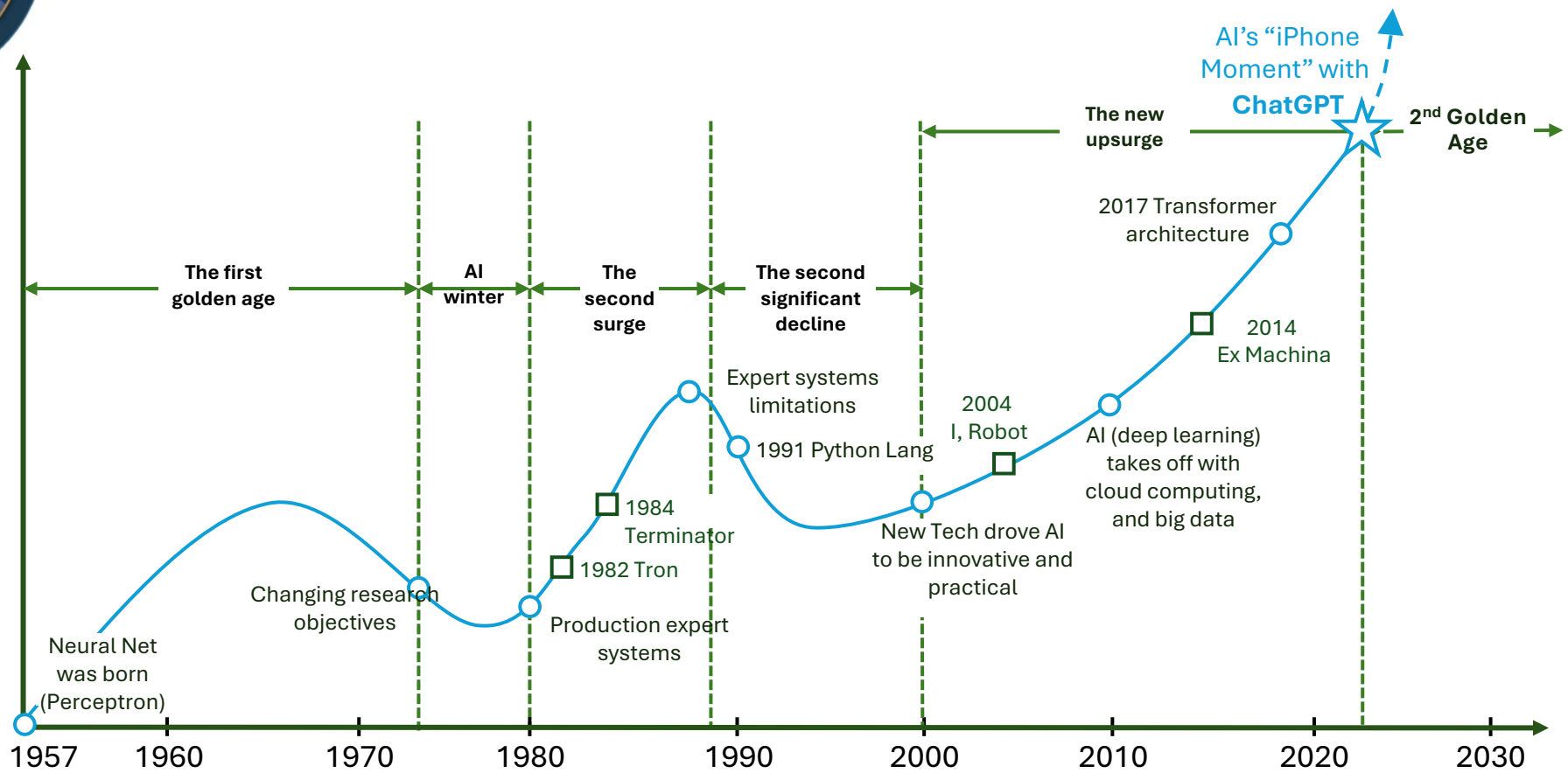


Figure adapted from the "Reflection of the development history of AI," published in the article "Artificial Intelligence in Product Lifecycle Management" by Wang et al. in 2021

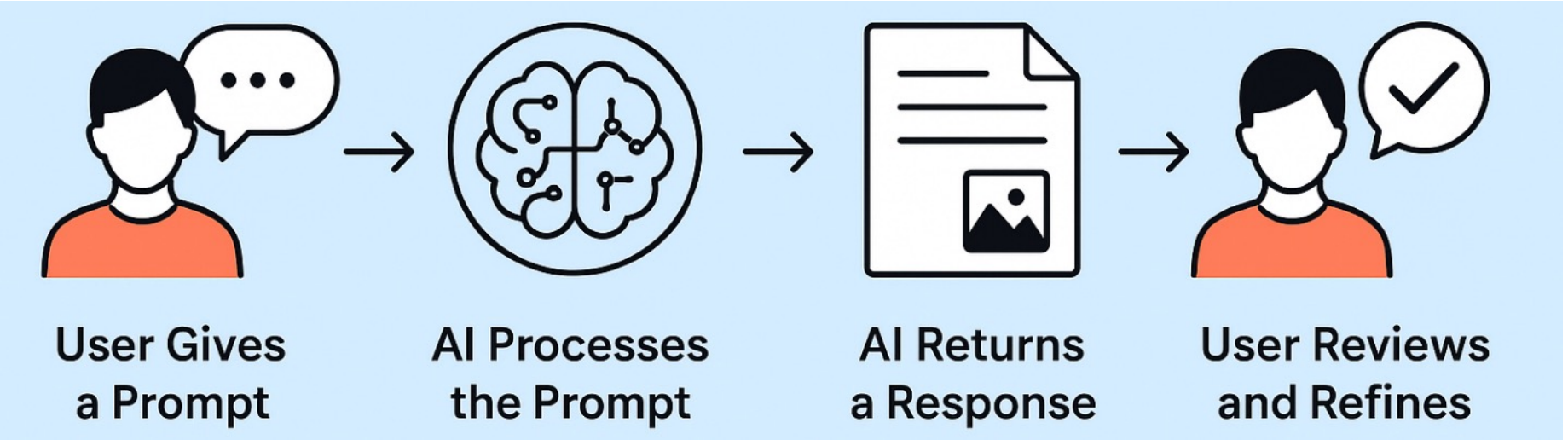
"Google's DeepMind AI Predicts 3D Structure of Nearly Every Protein Known to Science" CNET

"AI to stop water pollution before it happens" BBC

"How ChatGPT is changing the job hiring process, from the HR department to coders" CNBC

"Surgeons successfully restore touch and movement in quadriplegic man using AI brain implants" Euronews

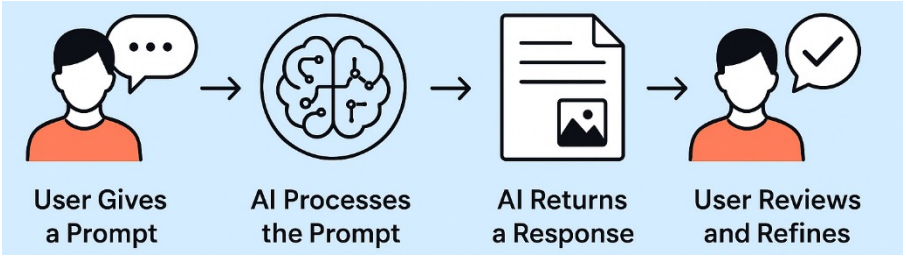
The Basics


User Gives a Prompt
AI Processes the Prompt
AI Returns a Response
User Reviews and Refines

The Basics

In a public AI model, what happens to your data?

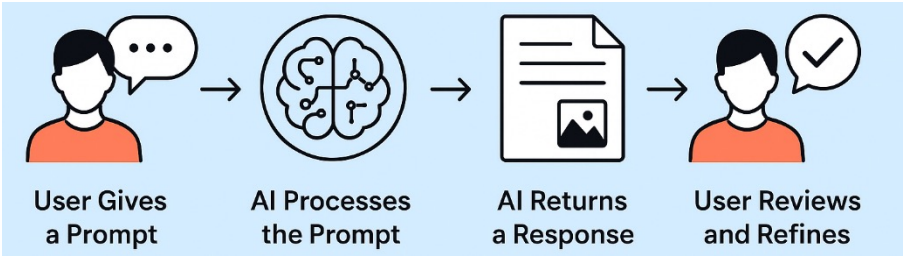
- Your prompt is sent to the AI provider's servers for processing.
- Your input may be stored and used to improve the model.
- Your prompts and the responses given may be manually reviewed by humans.
- Your data may be retained for a period of time.
- Unless otherwise stated, **your data may not be private!**



The Basics

How is an Enterprise AI model different?

- **Zero data retention** means your prompt and responses are not stored or used for training.
- Your organization's data is logically **separated** and **encrypted**.
- Allows access controls so AI only accesses data the user is **authorized** to see.
- Interactions are logged for compliance and governance **audits**.
- AI can be **grounded** in your organization's private data to generate more relevant, accurate responses.



Generative AI Approach Maturity

Prompt Engineering

Maturity: Foundational

Value: Fast, low-cost insights; easy adoption; early “wow” factor.

How it works: User types a question into an LLM and the LLM returns an answer.

Example: Commercial tooling to summarize meetings, draft an email, or support brainstorming

Key Considerations: Train users on effective prompting; establish style guidelines & basic governance.

RAG

Maturity: Emerging → Optimized

Value: Accurate, context-aware answers; reduces “hallucinations”; ensures compliance; speeds decisions.

How it works: System searches internal data (vector DB), feeds relevant context to LLM, returns company-specific answer.

Example: RAG solution connects multiple data sources in an enterprise environment for easy access to data for business users.

Key Considerations: Build/maintain vector index; define data access policies & refresh cadence; pilot in one department before scaling.

Agentic AI

Maturity: Optimized → Transformative

Value: End-to-end automation; significant time savings; frees staff for high-value work; scalable.

How it works: AI agent plans, has memory, uses tools (APIs/apps), and executes multi-step tasks autonomously.

Example: An enterprise helpdesk dynamically routes tickets and gives the user a first pass of resolution.

Key Considerations: Invest in an orchestration platform (agents, tools, memory); implement strong access controls & kill switches; start with human-in-the-loop pilots.

Risks associated with AI

“New AI systems collide with copyright law” *BBC*

“Hollywood's Fight: How Much AI Is Too Much?” *Wall Street Journal*

“Google hit with lawsuit alleging it stole data from millions of users to train its AI tools” *CNN*

“Cyber Attackers Can Disable AI Systems By 'Data Poisoning': Google AI Expert” *NDTV*

“OpenAI's ChatGPT chatbot blocked in Italy over privacy concerns” *Euronews*

“AI-enhanced images a ‘threat to democratic processes” *The Guardian*

RISKS INCLUDE

Privacy infringement	Exacerbating existing inequalities	Cultural biases
Abuse of AI to cause harm	Security vulnerabilities & data poisoning	Intellectual property issues, Plagiarism
Lack of accountability	Loss of trust, reputation & money	

Privacy Meets Intelligence

Safeguards in Modern AI

1 Data Privacy & Compliance

Data **anonymization**, **masking**, and **synthetic data generation** help prevent exposure of data.

2 Zero Data Retention

No **storage** or **reuse** of user prompts or outputs.

3 Access Controls and Encryption

Data is encrypted **at rest and in transit** and role-based access control allows only authorized users access to sensitive data.

4 Threat Detection & Response

Tools to detect **risky behavior**, **prompt injections**, and **data oversharing**.

5 Prompt Filtering & Output Monitoring

Runtime filters catch prompts that might elicit **sensitive** or **harmful** responses. Output monitoring tools score responses for **toxicity**, **bias**, and **hallucinations**.

6 Secure AI Frameworks

Frameworks offer structured approaches to securing AI systems, including risk assessments, audits, and implementing best practices.

Key Ethical Principles for Responsible Use of Data

- **Inclusivity and Relevancy**: Ensure data used for training and outputs is validated by the business and validated for inclusiveness and statistical relevancy to the AI problem statement. Continue to monitor AI outputs for relevancy.

- **Bias Protection**: Ensure AI models are addressing algorithmic limitations such as bias and variance.

- **Secure Data Pipelines**: Ensure any data ingested by AI uses secure data pipelines with sensitive data protections implemented to manage personally identifiable information (PII), IP and other sensitive or proprietary information.

- **De-identification**: Restrict access to any PII through de-identification, masking and analysis of the risk of re-identification as part of the data ingestion processes. These transformations should be implemented prior to access by any AI tools or models.

"Guardrails for data protection in the age of GenAI" published on cgi.com 2023

Key Ethical Principles for Responsible Use of Data

- **Cloud Security**: Ensure that data and information storage and data movement are secured within the cloud tenancy and apply access restrictions appropriate to the user level.

- **API Encryption**: Use secure and encrypted application programming interface (API) calls for AI access to any information or data retained on premises and apply de-identification to on-premises data, where appropriate.

- **SME Validation**: Ensure interpretation of AI responses and outputs is conducted with business and data subject matter experts to validate accuracy and relevancy to business operations.

- **Transparency**: Provide insight on the source of information in the responses as part of the AI outputs through citations and links to source information or documents.

"Guardrails for data protection in the age of GenAI" published on cgi.com 2023

Digital Forensics and AI

Security & Legal Ethics

Four Basic ABA Model Rules that Govern

Rule 1.1	**Competence**
Rule 1.4	Communications
Rule 1.6	Duty of **Confidentiality**
Rule 5.1, 5.2, 5.3	Lawyer & Nonlawyer Associations

The **"Big Two"** in **Artificial Intelligence** and **Cybersecurity**

Begin Your Journey Toward **Competence** to Keep *Office* Data, Documents, and Communication **Confidential**

40 States Have Adopted Revised Rule 1.1

"To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, *including the benefits and risks associated with relevant technology*"

AI & Legal Ethics

AMERICAN BAR ASSOCIATION

STANDING COMMITTEE ON ETHICS AND PROFESSIONAL RESPONSIBILITY

Formal Opinion 512 **July 29, 2024**

Generative Artificial Intelligence Tools

To ensure clients are protected, lawyers using generative artificial intelligence tools must fully consider their applicable ethical obligations, including their duties to provide competent legal representation, to protect client information, to communicate with clients, to supervise their employees and agents, to advance only meritorious claims and contentions, to ensure candor toward the tribunal, and to charge reasonable fees.

AI Contract Compliance

"No Confidential Information may be uploaded to, processed by, or disclosed to any publicly available AI/ML system, model, or dataset without prior written consent."

AI-specific clauses are starting to appear in some contracts as legal teams and AI law specialists are recommending clauses that:

- Ban feeding confidential data into public models without written consent.
- Require proof that approved tools will not train on the data.
- Bind contractors and sub-processors to same rules

Steven Schwartz

Steven Schwartz – NY Personal Injury Lawyer

Drafted court filing relying on AI chatbot ChatGPT

Motion had **made-up cases, rulings and quotes**

Lawyer filed motion after ChatGPT assured him:

> “The cases I provided are real and can be found in reputable legal databases”

(They were not, and can not)

Canadian Lawyer Chong Ke

Turned to ChatGPT when her client wanted to know if he could take his children on an overseas trip in the midst of a custody dispute.

To make her argument, Ke cited precedent from two court cases that ChatGPT had supplied—both of which were completely made up.

When it was all said and done, Ke had to pay the court costs it took for the opposing counsel to research the nonexistent cases.

Caveat Emptor – Due Diligence

Falsehoods, Lies -> AI Hallucinations

AI Foundation LLMs are trained on:

Entire Internet, books, magazines, blogs, etc.

Also, legal documents and evidence

AI generates language based on lawyers' prompts

But, crucially, does not Shepardize it!

Lawyer’s obligation to Shepardize, KeyCite, etc.

Check your citations!!

Expert Prompt Engineer's Advice

Large Language Models:

Sometimes Wrong

Never in Doubt

Always Confident

Never Say, "I Don't Know"

AI Confidentiality & Security

Public Documents are Safe

- Statutes, Rules, Regs, Published Opinions

Generative AI Exposes Sensitive, Privileged, Client Data by Sharing with LLMs to Train

- Exposes User Prompts that Guide AI Tool
- Exposes Legal Documents Uploaded to AI Tool
- Too Much Account Access via Permissions

Confidentiality Breaches of ACP and WPD and NDAs

AI Confidentiality & Security

Know Your AI Tool

Choose AI Legal Drafting Tools Carefully

- Confidentiality
- No Sharing with LLM Training Data

Examples: Westlaw AI-Assisted Research, Westlaw Quick Check, Casetext CoCounsel, vLex Vincent AI, Lexis + AI, BriefCatch

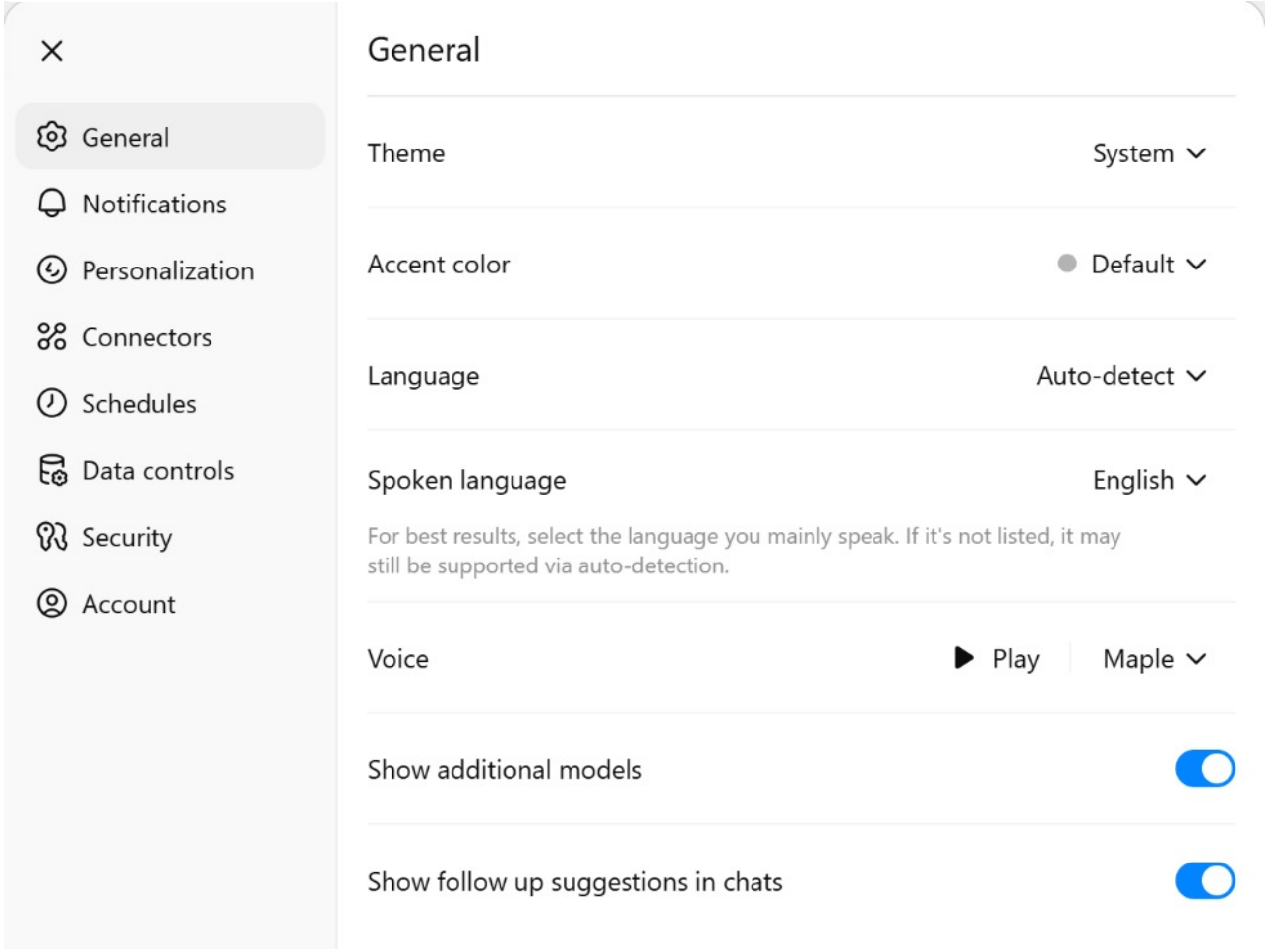
AI Confidentiality

Opt Out of Sharing Training Data with LLMs (e.g. ChatGPT, Gemini, Claude)

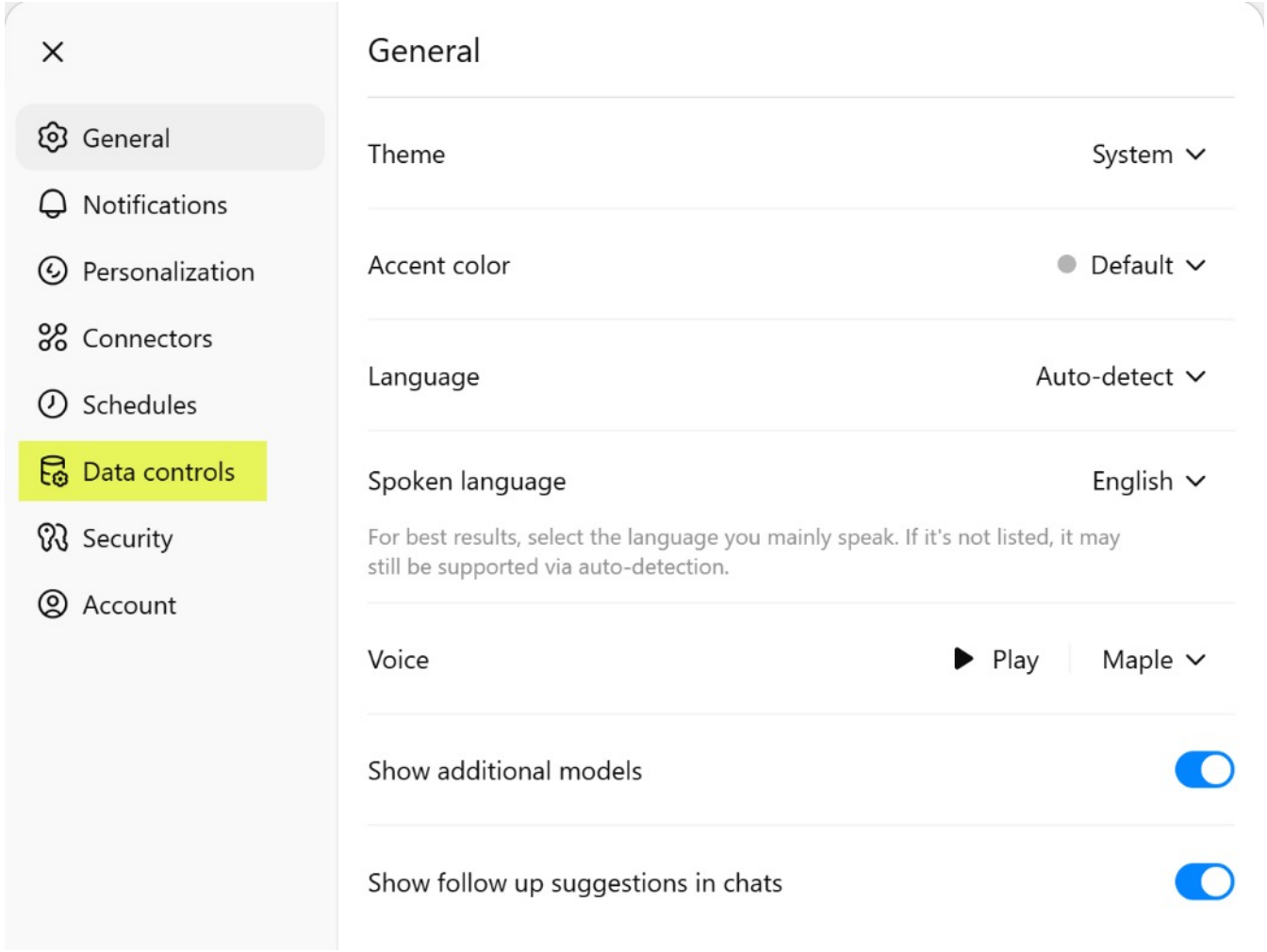
Eliminate Data Leakage

ChatGPT Plus Confidentiality

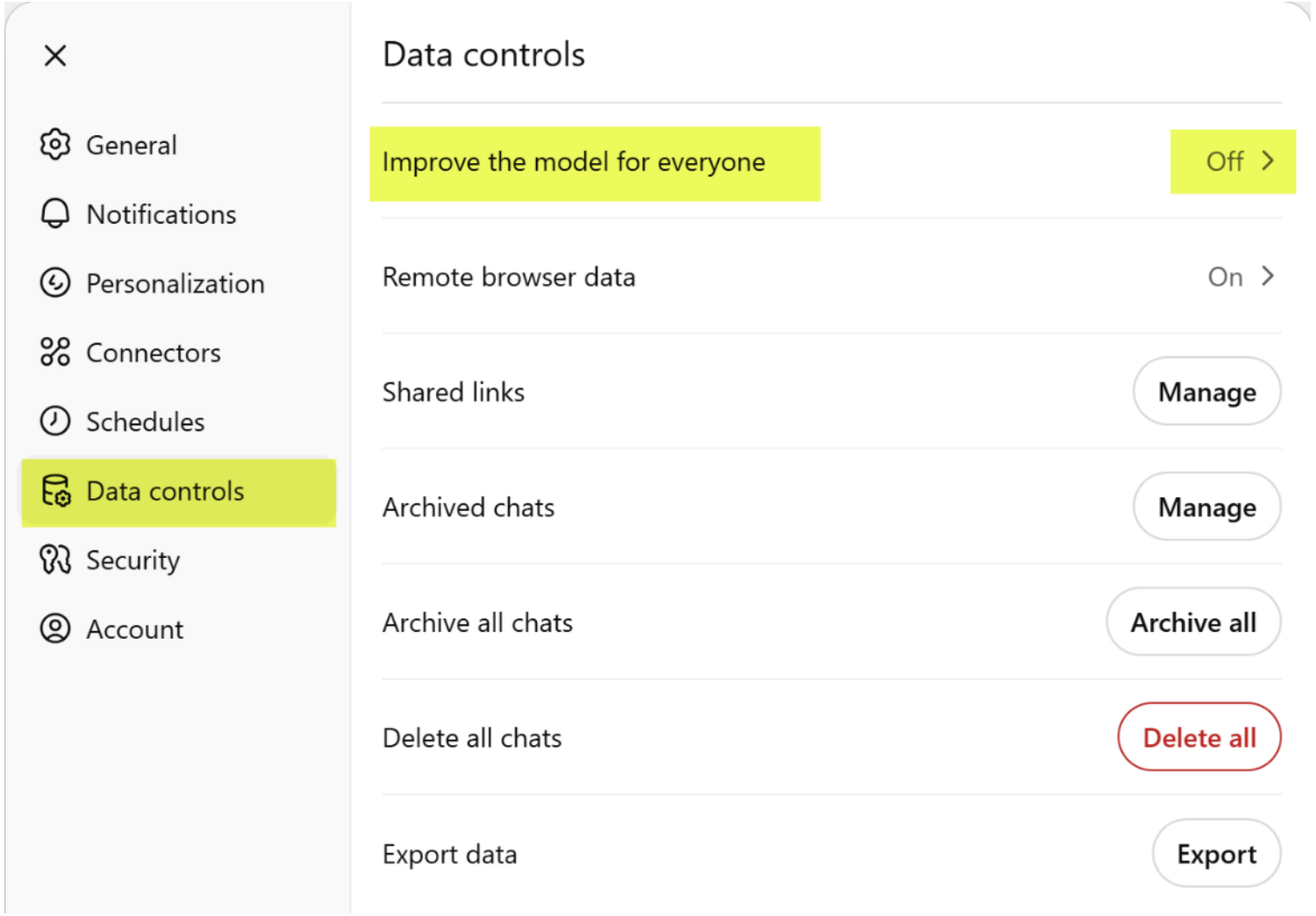
Settings



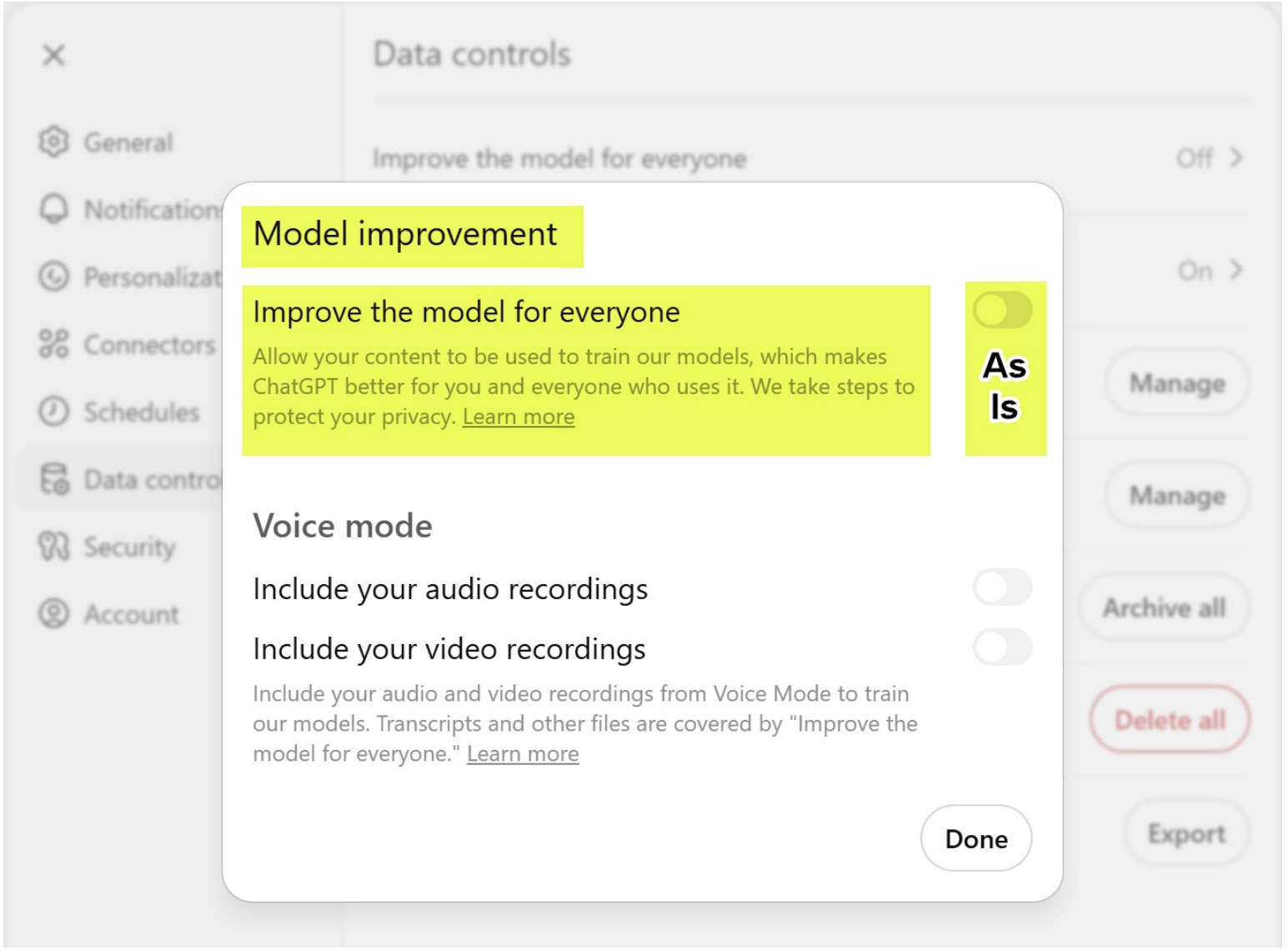
ChatGPT Plus Confidentiality



ChatGPT Plus Confidentiality



ChatGPT Plus Confidentiality



Gemini Confidentiality

Your Gemini Apps Activity

Gemini Apps give you direct access to Google AI. Your chats are saved in your account for up to 72 hours, whether Gemini Apps Activity is on or off. Google uses this data to provide the service, maintain its safety and security, and process any feedback you choose to provide.

Gemini Apps Activity
Saving activity

Turn off

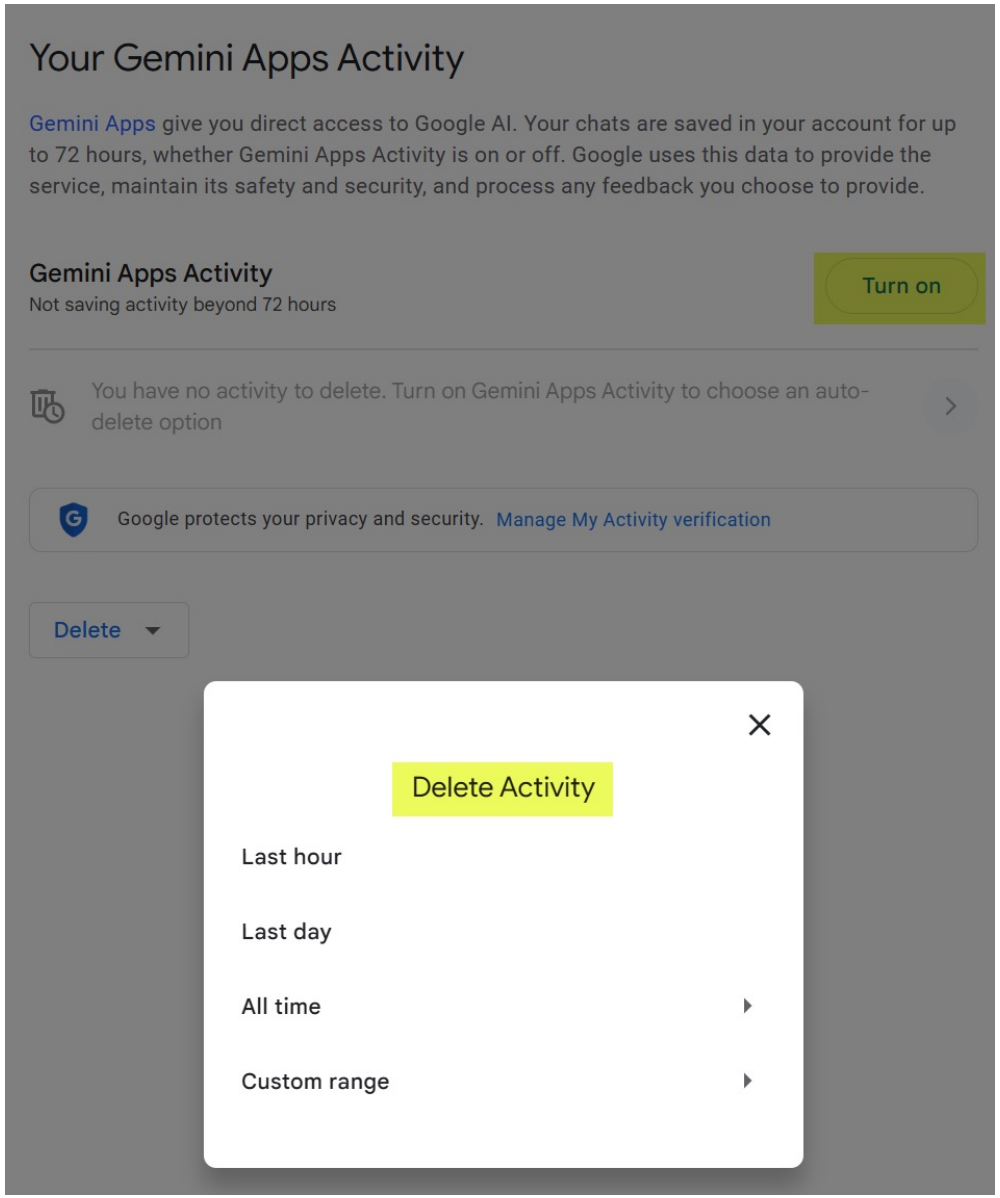
Turn "Activity" Off!

Deleting activity older than 18 months

Google protects your privacy and security. Manage My Activity verification

Delete

Gemini Confidentiality



Gemini Workspace Confidentiality

Google Workspace Admin Help

Describe your issue

Business / Enterprise | Education

- What is the Gemini app?
- Who can use the Gemini app?
- What if I don't have Google Workspace today? What if I don't have the right edition?
- How does Gemini for Google Workspace protect my data?

When Google Workspace commercial customers adopt Gemini for Google Workspace, they get the same robust data protection and security standards that come with all Google Workspace services, with specific protections for businesses, education, and public-sector customers:

- Users with a Gemini for Google Workspace license get enterprise-grade data protections when they use Gemini app. Submissions aren't used to train models and are never reviewed by humans. Learn more about how Gemini app works for users who don't have a license.
- Your interactions with Gemini for Google Workspace stay within your organization. Gemini for Google Workspace stores any prompts or generated content alongside your Workspace content and does not share them outside your organization.
- Your existing Google Workspace protections are automatically applied. Gemini for Google Workspace brings the same enterprise-grade security as the rest of Google Workspace, automatically applying your organization's existing controls and data handling practices, such as data-regions policies and Data Loss Prevention.
- Your content is not used for any other customers. None of your content is used for model training outside of your domain without permission.

Gemini Workspace Confidentiality

Gemini app

Users

Groups

Organizational Units

Showing settings for users in Test Org

Gemini conversation history

Gemini conversation history
Applied at 'Test Org'

Gemini conversation history controls allow admins to control if conversations with Gemini are saved and for how long before they are automatically deleted. Disabling Gemini conversation history will not affect previous conversation history which is subject to the auto-delete setting below.

Gemini conversation history
Gemini conversation history allows users to return to past conversations

Enable Gemini conversation history
Uncheck to turn Gemini history off

Conversation retention
Choose when conversations should be deleted from a user's history

18 months

Claude Sonnet 4 – Free Plan

Settings

Profile

Account

Privacy

Billing

Features

Connectors

Claude Code

Full name

John Carney

What should Claude call you?

John

What best describes your work?

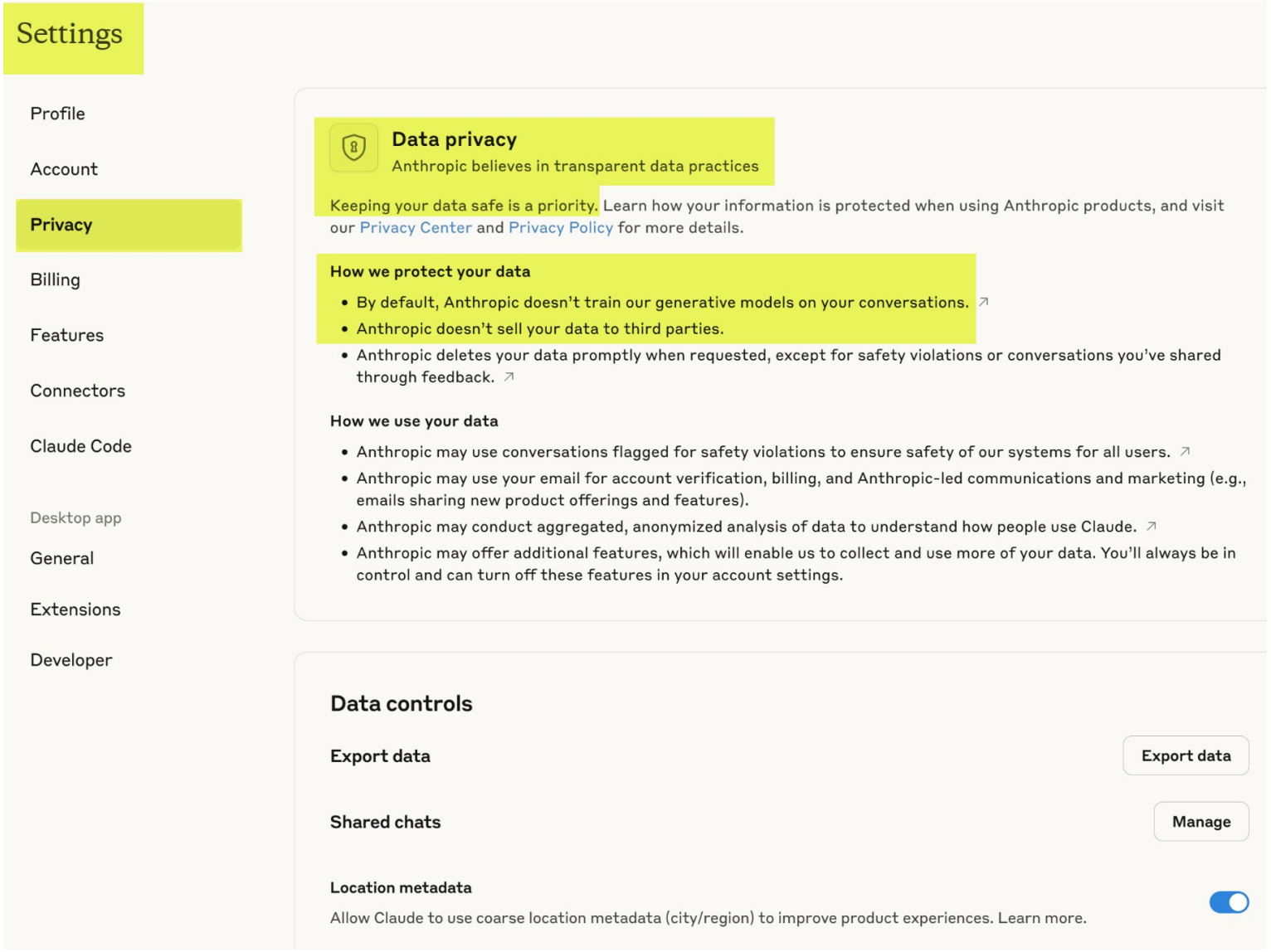
Select your work function

What personal preferences should Claude consider in responses? BETA

Your preferences will apply to all conversations, within Anthropic's guidelines. Learn about preferences

e.g. when learning new concepts, I find analogies particularly helpful

Claude Sonnet 4 – Free Plan

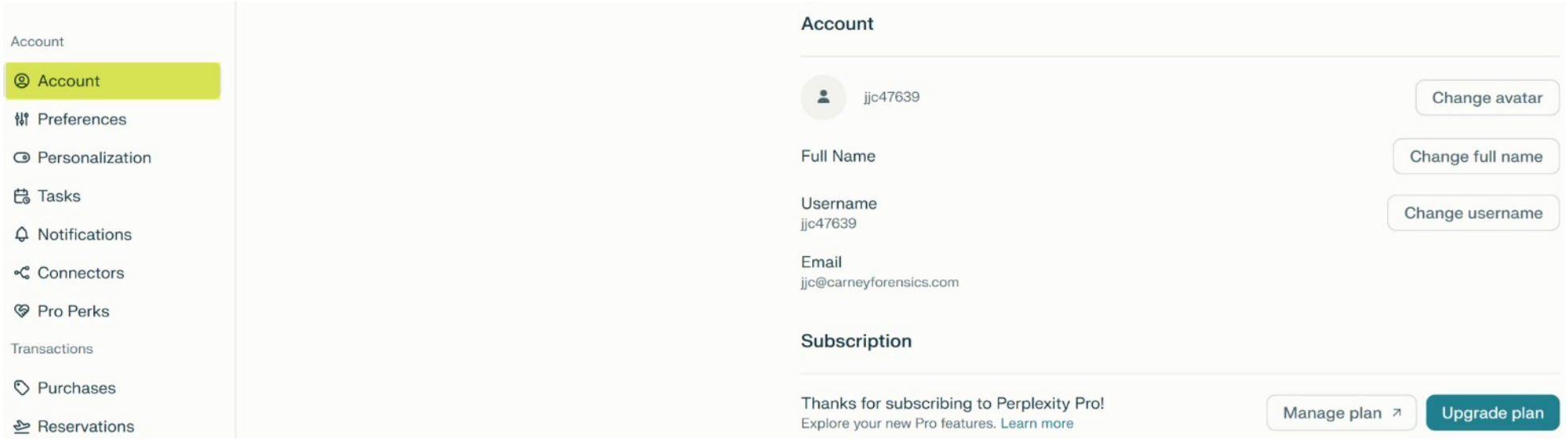


AI Confidentiality

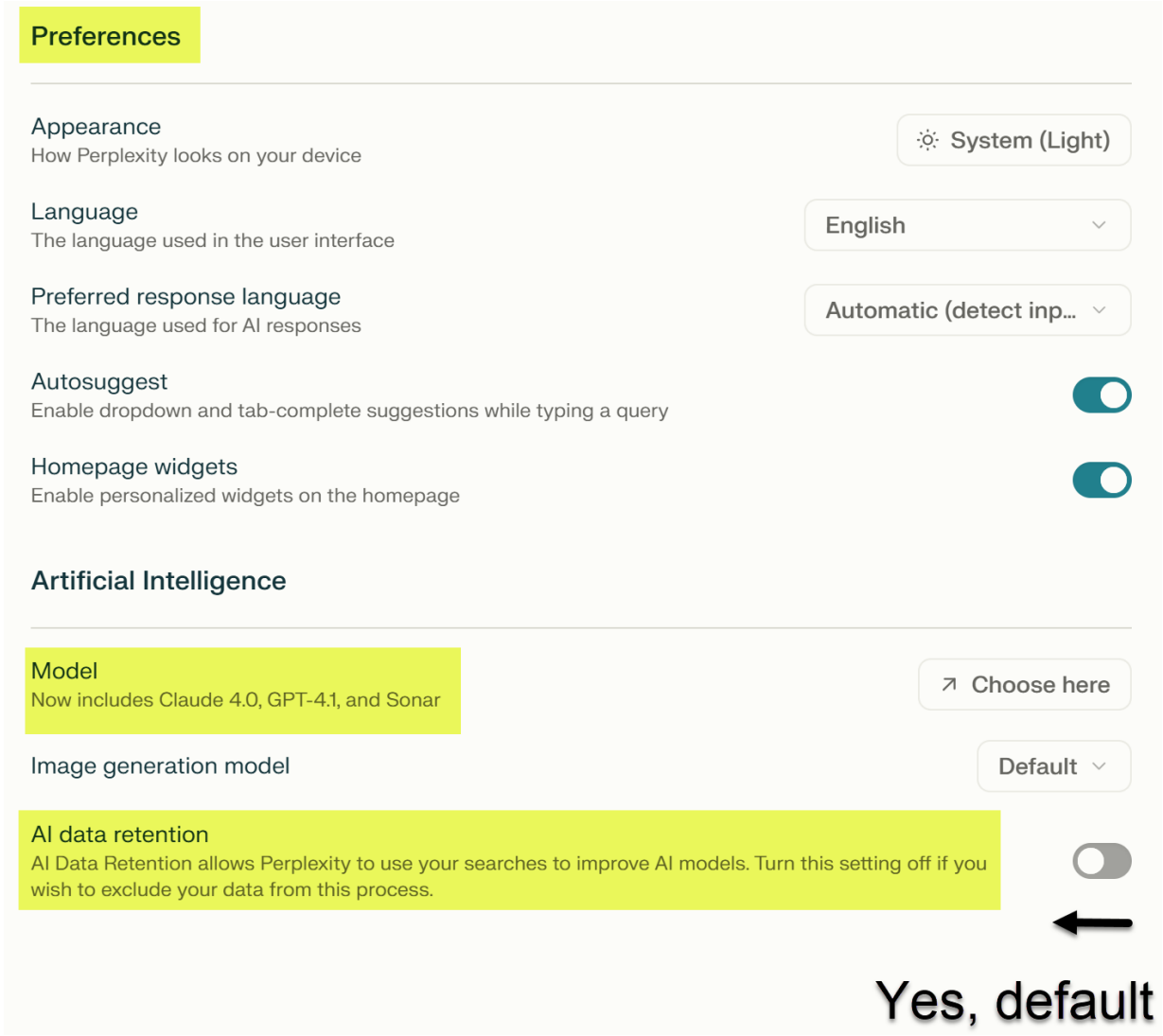
Opt Out of Each LLM for AI Tools that use multiple LLMs (e.g. Perplexity)

Eliminate Data Leakage

Perplexity Confidentiality



Perplexity Confidentiality



AI Confidentiality

Use Professional / Paid Version of AI Tool

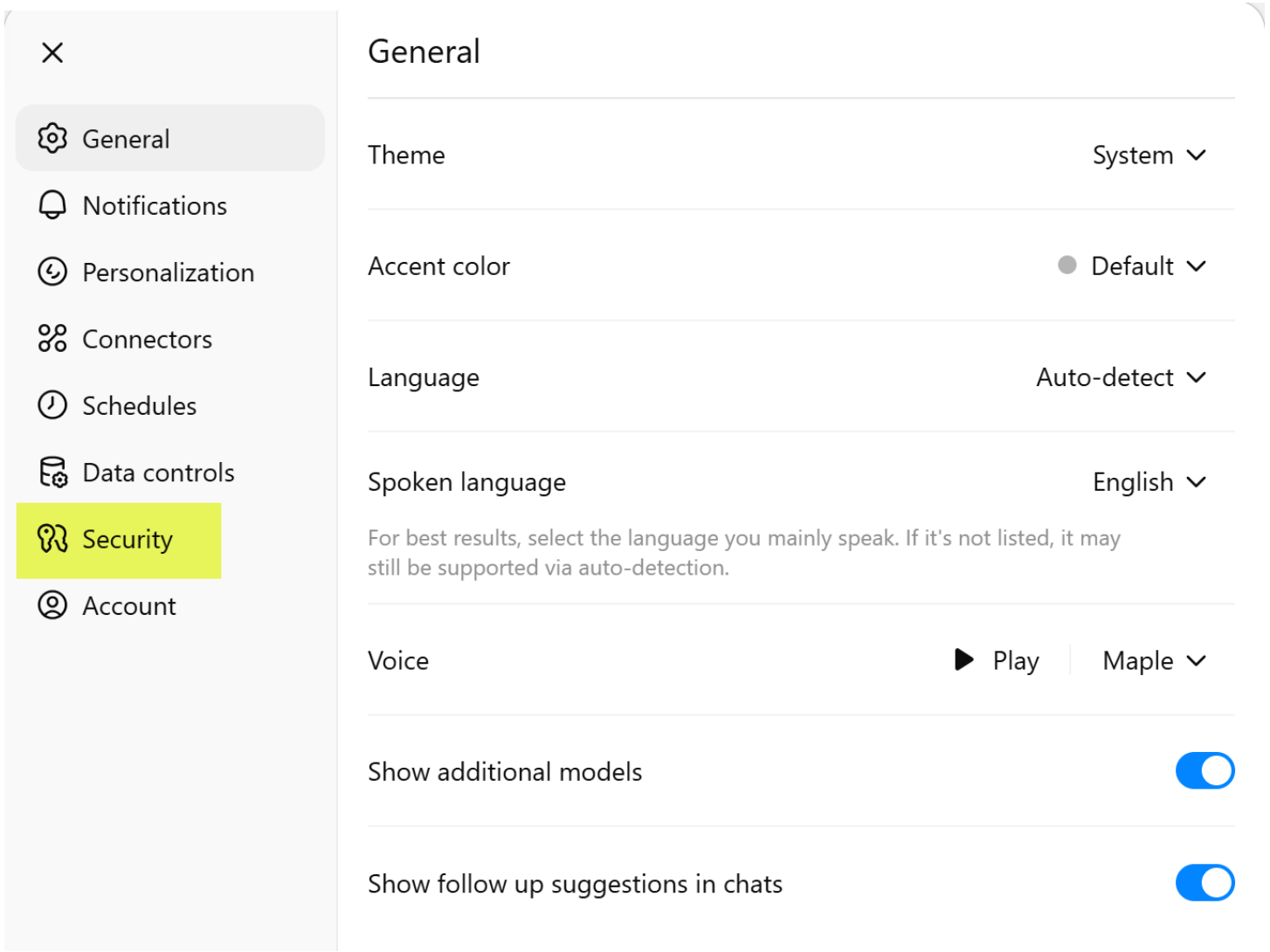
- Not the Free Version
- Pro / Paid Version has Settings
- Settings for Opting out of LLMs

AI Security

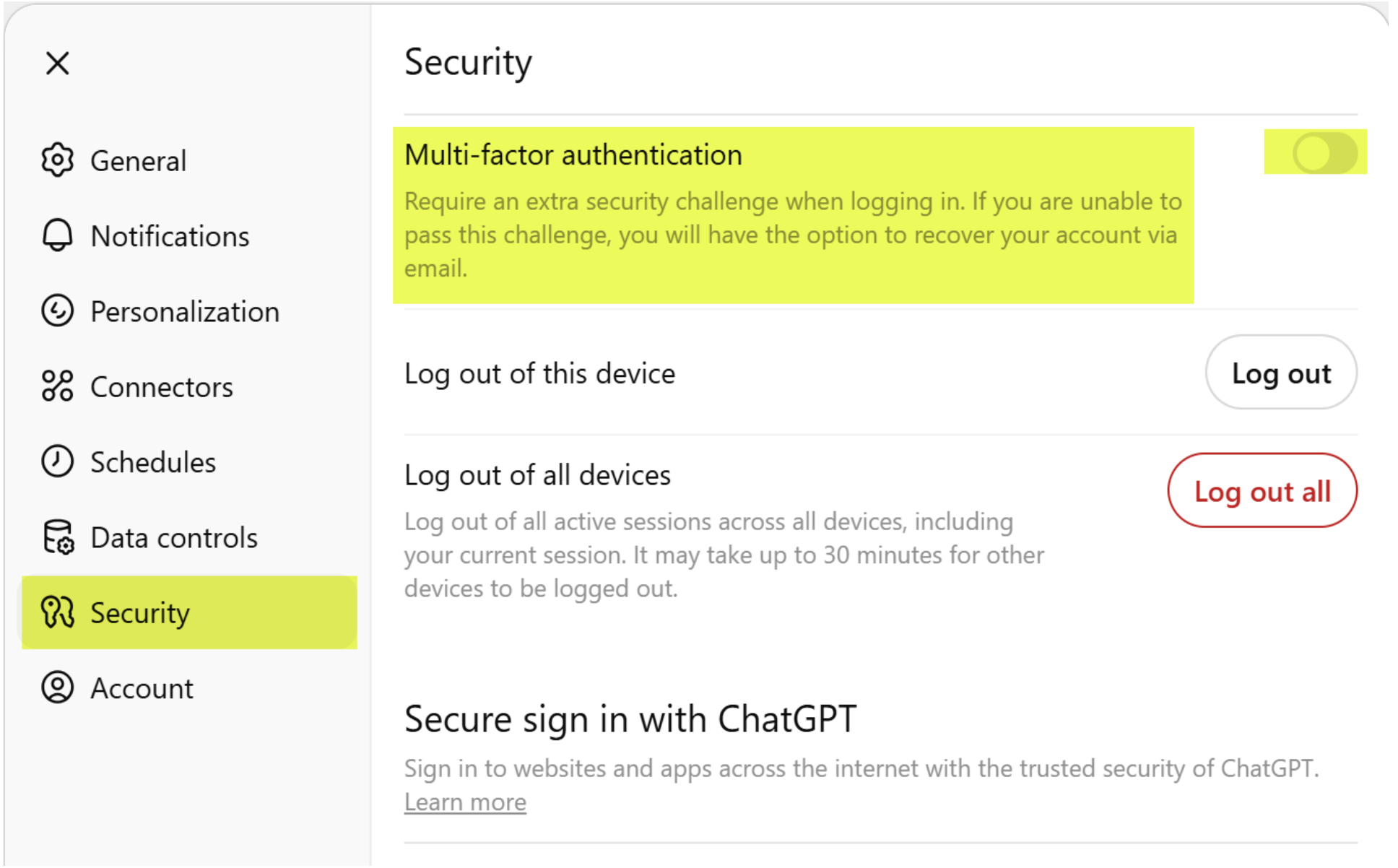
- Apply Modern Cybersecurity Practices
  - Strong Credentials
    - Complex Passcode, Not Reused
  - Apply 2FA Protection to AI Account
  - Use Encryption for Prompts, Uploaded Documents, AI Threads

ChatGPT Plus Security

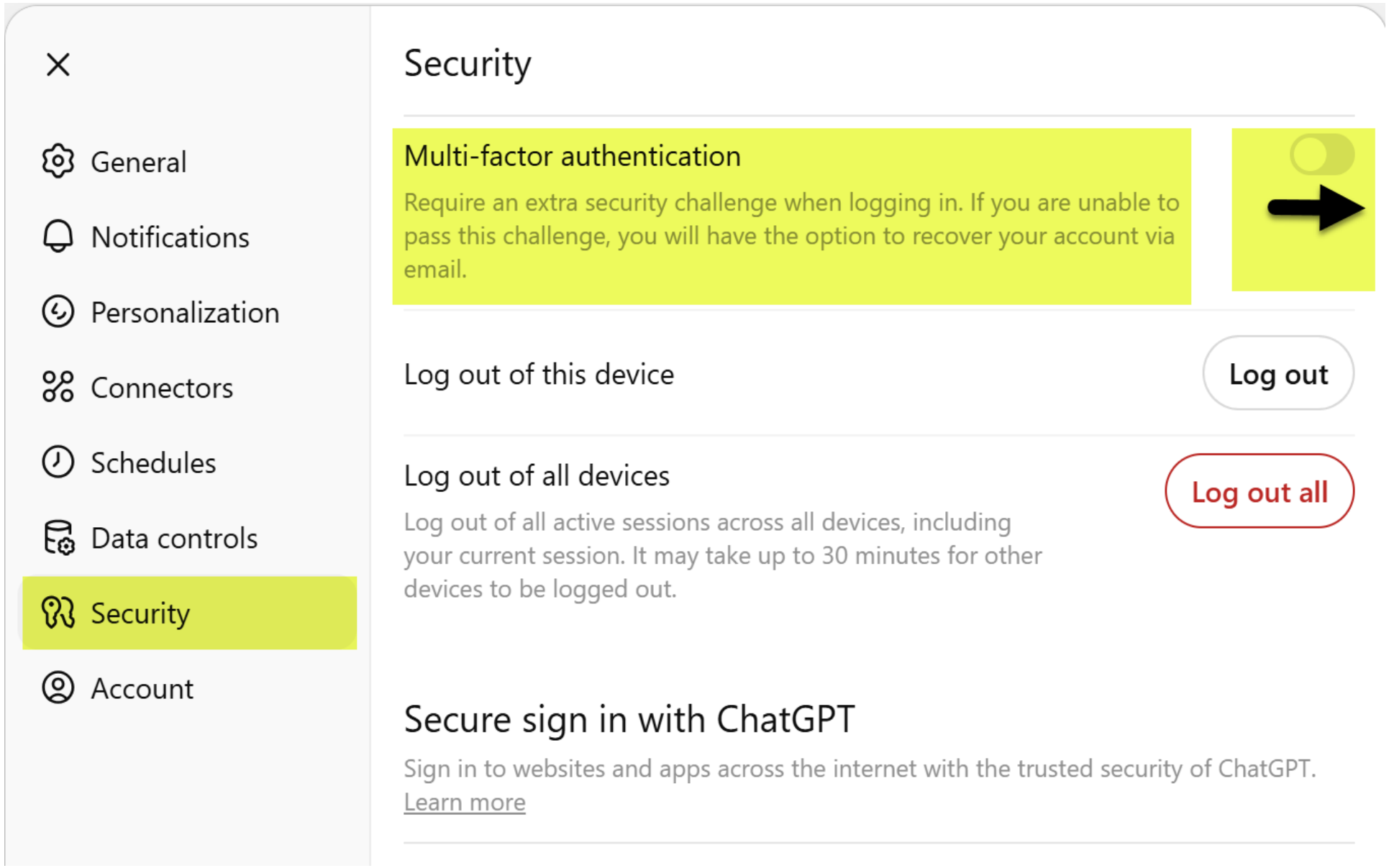
Settings



ChatGPT Plus Security



ChatGPT Plus Security



Main Takeaways


Q&A

THANK YOU!


WESTERN INTERGOVERNMENTAL CHILD SUPPORT ENGAGEMENT COUNCIL
Navigating the Currents of Change
ST. PAUL, MN | SEPT. 8-12, 2025

Contact Information:

John Carney: jjc@carneyforensics.com

Carl Medley: carlmedley@eworldes.com

Zachary Rasmussen: zachary.rasmussen@cgi.com