A Guide to the Safe and Secure Use of Artificial Intelligence in Research

1. Introduction

Artificial Intelligence (AI) presents transformative opportunities for research across all disciplines at UC Riverside. From analyzing vast datasets to generating novel hypotheses, AI tools can accelerate discovery and innovation. However, the use of AI also introduces new challenges related to data security, privacy, and research integrity.

This document provides a comprehensive guide for all UCR researchers on the approved use of AI models and infrastructure. It is designed to help you navigate the complexities of using AI in your work while ensuring compliance with University of California policies, particularly the **UC Information Security Policy IS-3**, and protecting the confidentiality and integrity of your research data.

2. Understanding Data Protection Levels (UC IS-3)

The University of California classifies institutional information and IT resources into four Protection Levels (P1-P4) and four Availability Levels (A1-A4). The Protection Level is determined by the potential impact of a data breach on individuals or the university. All researchers must classify their data according to these levels to determine the appropriate security controls.

Here is a summary of the Protection Levels:

Level	Name	Description	Examples
P1	**Public**	Information intended for public access.	Public-facing websites, press releases, published research, course catalogs.
P2	**Internal**	Information not generally intended for public access and may be accessed on a "need-to-know"	Staff and academic personnel records not containing P4 information, unpublished research

		basis.	data not subject to specific protections.
P3	**Confidential**	Information that is confidential and protected by law or contract. Unauthorized disclosure could have a significant adverse impact on the university or individuals.	Student records (FERPA), non-public research data with contractual protections, personally identifiable research subject information.
P4	**Highly Confidential**	Information whose unauthorized disclosure could result in severe harm, including significant financial loss, legal liability, or reputational damage.	Personally Identifiable Health Information (PHI) under HIPAA, credit card and payment information (PCI-DSS), Social Security numbers, export-controlled research data.

For a complete list of data classifications, please refer to the official UC Data Classification Standard.

3. Approved AI Models

Researchers have access to a wide range of AI models. The appropriate model for your research will depend on your specific needs, the nature of your data, and the approved infrastructure you are using.

3.1. Closed-Source Models

- **Gemini (via Google Cloud):** Google's Gemini family of models is approved for use within the UCR-managed Google Cloud environment ("Ursa Major"). These models offer state-of-the-art capabilities for a variety of tasks.
  - **Data Privacy:** When used within the UCR Google Cloud environment, your data is not used to train Google's models, and your prompts and data remain within the UCR tenant.

3.2. Open-Source Models

- **Via Vertex AI Model Garden on Google Cloud:** UCR's "Ursa Major" Google

Cloud environment provides access to a curated list of open-source models through the Vertex AI Model Garden. This is the recommended way to use open-source models for most research.
- **On-Premise (HPCC or a private system):** Researchers can run open-source models on the UCR High-Performance Computing Center (HPCC) or on their own secure, departmentally-managed systems.

4. What is Vertex AI?

Vertex AI is Google Cloud's unified machine learning (ML) platform, designed to help researchers and developers build, deploy, and scale ML models more efficiently. It provides a comprehensive suite of tools that support the entire ML lifecycle, from data preparation to model deployment and management, all within a single interface. For UCR researchers, it is the primary gateway to accessing advanced AI capabilities on Google Cloud.

Key Capabilities of Vertex AI:

- **Unified Environment:** It brings together all of Google's cloud services for building ML under one roof, eliminating the need to piece together separate services.
- **Data Preparation & Labeling:** Provides tools to ingest, analyze, and prepare your datasets for training, including data labeling services to create high-quality training data.
- **AutoML & Custom Training:** Researchers can use AutoML to automatically train high-quality models with minimal effort and ML expertise, or use Custom Training to have full control over the model architecture and training process using popular frameworks like TensorFlow, PyTorch, and Scikit-learn.
- **Model Garden:** As mentioned, this is a central repository of pre-trained and open-source models (including from Hugging Face) that can be easily deployed or fine-tuned for specific research tasks.
- **MLOps Tools:** Vertex AI includes a robust set of MLOps (Machine Learning Operations) features, such as pipelines for automating workflows, a model registry for versioning and management, and monitoring tools to track model performance and detect drift. This helps ensure that your research is reproducible, scalable, and manageable over time.
- **Generative AI Studio:** A dedicated environment within Vertex AI for prototyping and customizing generative AI models like Gemini. Researchers can design prompts, tune models with their own data, and deploy them for use in applications.

By leveraging Vertex AI, UCR researchers can significantly accelerate their research workflows, moving from idea to production-ready model faster and with greater ease.

5. Approved Infrastructure and Data Usage

The infrastructure you use to run AI models is directly tied to the level of data protection required for your research. The following table outlines the approved configurations and the corresponding data protection levels they are cleared to handle.

Infrastructure	Approved Models	P1 Data	P2 Data	P3 Data	P4 Data
UCR "Ursa Major" Google Cloud	Gemini, Vertex AI Model Garden	✅	✅	✅	⚠️
UCR High-Performance Computing Center (HPCC)	Open-Source Models	✅	✅	✅	❌
Researcher/Department-Managed Systems	Open-Source Models	✅	✅	⚠️	❌

⚠️ **Important Considerations:**

- **P4 Data on Google Cloud:** Using P4 data with AI models on Google Cloud requires a formal risk assessment and the implementation of specific security controls, such as data de-identification and the use of a secure, compliant environment. **You must consult with the UCR Information Security Office before using P4 data with any AI model.**
- **P3 Data on Researcher Systems:** Using P3 data on a researcher or department-managed system requires a robust security plan that meets the requirements of IS-3. This includes, but is not limited to, data encryption, access controls, and regular security patching. **Consult with your department's IT lead and the Information Security Office.**
- **Export Control:** Research data subject to export control regulations has specific

handling requirements. If your research involves such data, you must consult with the Office of Research Integrity.

6. Safe Storage of Research Data

Properly storing your data is as important as choosing the right AI model and infrastructure. The appropriate storage solution depends on the Protection Level of your data.

Protection Level	Approved Storage Solutions	Key Considerations
P1 / P2	UCR Google Drive, UCR OneDrive, HPCC Storage, Departmental Servers	These solutions are suitable for data that is public or for internal use.
P3	UCR Google Drive (with appropriate sharing restrictions), HPCC Storage, Secure Departmental Servers	Requires strict access controls. Data should only be shared on a "need-to-know" basis. Encryption is mandatory.
P4	**Must be approved by the Information Security Office.** Typically requires a dedicated, secure environment like a secured enclave within Google Cloud or a physically secured, encrypted server.	**Do not store P4 data in standard cloud storage or on unencrypted devices.** A formal Data Security Plan and risk assessment are required.

General Best Practices for All Data Levels:

- **Encryption:** All devices used to store or access P2, P3, and P4 data (laptops, external drives) must be encrypted.
- **Access Control:** Use the principle of least privilege. Only grant access to individuals who absolutely require it for their research duties.
- **Data Minimization:** Collect and retain only the data that is essential for your research.
- **Secure Deletion:** When data is no longer needed, ensure it is disposed of securely according to university policy.

7. Getting Started

- **For access to UCR's "Ursa Major" Google Cloud environment**, please contact Research Computing at research-computing@ucr.edu.

- **For access to the HPCC**, please visit the HPCC website for information on getting an account.
- **For questions about data classification and security**, please contact the UCR Information Security Office.

8. Conclusion

AI offers powerful tools to enhance our research capabilities. By following these guidelines, we can leverage these tools responsibly, ensuring the security and integrity of our research and the data we are entrusted to protect. **Information Technology Solutions (ITS)** is committed to supporting the research community in this endeavor. We encourage you to reach out with any questions or for assistance in navigating the use of AI in your work.

9. Appendix: Available Models in Vertex AI Model Garden

This list is a sample of the models available and is subject to change. In addition to the models listed below, Vertex AI Model Garden provides access to thousands of the most popular models from ***Hugging Face*** *for easy installation and use. For the most current list, please consult the Vertex AI Model Garden directly within the Google Cloud console.*

9.1. Foundation Models

Google Models

- **Gemini 2.5 Flash-Lite:** Most balanced Gemini model for low latency use cases.
- **Gemini 2.5 Pro:** Strongest model quality, especially for code & complex prompts.
- **Gemini 2.5 Flash:** Best for balancing reasoning and speed.
- **Gemini 2.0 Flash-Lite:** Our cost-effective Gemini model to support high throughput.
- **Gemini 2.0 Flash:** Workhorse model for all daily tasks.
- **Gemini 1.5 Pro:** Created to be multimodal (text, images, videos) and to scale across a wide range of tasks.
- **Gemini 1.5 Flash:** The best performing Gemini model with features for a wide range of tasks.
- **Gemini 1.0 Pro & Pro Vision:** Designed to balance quality, performance, and cost for various tasks including multimodal (text, images, code) applications.
- **Gemma Family (Gemma 3n, MedGemma, Gemma 3, ShieldGemma 2, Gemma 2, PaliGemma, CodeGemma, T5Gemma, TxGemma):** Lightweight, state-of-the-art open models from Google for various specialized tasks.
- **Imagen Family (4 Ultra, 4 Fast, 4, 2, Product Recontext, Virtual Try-On):** A

suite of models for high-quality image generation, editing, and customization.

- **Lyria 2 for Music Generation:** Generates high-quality instrumental music from text.
- **PaLM 2 (Chat Bison, Text Bison, Text Unicorn):** Models designed for natural conversation and single-turn instruction tasks.
- **Codey (Code Chat, Code Generation, Code Completion):** A family of models specialized for code-related assistance.
- **Chirp 2:** A multilingual model for speech-to-text transcription.
- **Health & Science Models (HeAR, Path Foundation, Derm Foundation, CXR Foundation):** Foundation models that produce embeddings for specialized health and science data (acoustic, pathology, dermatology, chest X-rays).
- **Other Google Models (WeatherNext, TimesFM, Embeddings for Text/Multimodal):** Specialized models for weather forecasting, time-series forecasting, and generating data embeddings.

Third-Party & Open-Source Models

- **Anthropic - Claude Family (Opus 4, Sonnet 4, 3.7 Sonnet, 3.5 Sonnet, 3 Haiku):** A range of powerful models from Anthropic, excelling in tasks from coding and research to fast, user-facing chatbots.
- **Meta - Llama Family (Llama 4, 3.1, 3.3, 3.2, 3, 2, Guard, Prompt Guard, Code Llama):** State-of-the-art open large language models from Meta, available as both APIs and for self-deployment.
- **Mistral (OCR, Codestral, Small 3.1, Large, Mixtral, 7B & Nemo):** A family of efficient and high-performing models from Mistral AI for various tasks including OCR and code generation.
- **AI21 Labs - Jamba Family (Large 1.6, 1.5 Large, 1.5 Mini):** Powerful models with very large context windows, optimized for long-form input, accuracy, and speed.
- **Microsoft - Phi Family (Phi-4, Phi-3):** Explore and build with Microsoft's Phi models.
- **Stability AI (Stable Diffusion XL Lightning, XL LCM, XL, v2.1, 4x-upscaler, Inpainting):** A suite of popular latent diffusion models for fast, high-fidelity text-to-image generation and editing.
- **Other Notable Models:**
  - **Image Generation:** Flux, HiDream-I1, Instant ID
  - **Speech & Audio:** Sesame CSM, MARS7, Dia-1.6B, Whisper Large
  - **Multimodal & Vision:** CogVideoX-2b, MaMMUT, LLaVA, OWL-ViT, CLIP, BLIP
  - **Specialized:** BioGPT (biomedical text), LayoutLM (document understanding), NLLB (translation for 200 languages)

9.2. Task-Specific & Pre-Built Models

Text & Language

- **Translation LLM:** The best performing translation model, fine-tuned from Gemini specifically for translating text between languages.
- **Text Translation:** Use Google's proven pre-trained text model to get text translations for 100+ languages.
- **Text Moderation:** Analyzes a document and returns a list of harmful and sensitive categories.
- **Syntax analysis:** Extracts linguistic information, breaking up text into sentences and tokens.
- **Entity sentiment analysis:** Identifies the prevailing emotional opinion of an entity within the text.
- **Sentiment analysis:** Determines the overall attitude (positive or negative) expressed within the text.
- **Content classification:** Analyzes text content and returns content categories for the content (supports over 1,000 categories).
- **Entity analysis:** Inspects text to identify and label persons, organizations, locations, events, products and more.

Code

- **Codestral:** A cutting-edge model specifically designed for code generation, including fill-in-the-middle and code completion.
- **Qodo-Embed-1-7B:** A suite of large-scale code embedding models for efficient code & text retrieval.

Video Analysis

- **Video Speech Transcription:** Transcribes speech in video files.
- **Video Text Detection:** Detects visible text in video files.
- **MoViNet Video Action Recognition/Clip Classification:** Efficient models for video classification tasks.
- **Bytetrack Multi-Object Tracking:** Detects, identifies, and tracks objects across video frames.
- **Person blur:** Masks or blurs a person's appearance in video.
- **PPE detector:** Identifies people and personal protective equipment (PPE).
- **Object detector:** Identifies and locates objects in video.
- **Person/vehicle detector:** Detects and counts people and vehicles in video.
- **Occupancy analytics:** Detects people and vehicles, plus zone detection, dwell time, and more.

Image Analysis

- **BiomedCLIP:** Zero-shot image classification with the BiomedCLIP biomedical vision-language foundation model.
- **Pic2Word Composed Image Retrieval:** A state-of-the-art image retrieval model.
- **Watermark detector:** Detects watermarks in an input image.
- **Text detector (Vision API):** Detects and extracts text from images using OCR.
- **Face detector (Vision API):** Detects multiple faces and provides bounding polygons and facial landmarks.
- **Content moderation (Vision):** Detects objectionable or unwanted content across predefined or custom labels.
- **Tag recognizer:** Extracts text in product and price tags.
- **Product recognizer:** Identifies products at the GTIN or UPC level.

Document AI

- **Form Parser:** Extracts key-value pairs, checkboxes, and tables from documents in over 200+ languages.
- **Document AI OCR processor:** Identifies and extracts text from documents in over 200 printed and 50 handwritten languages.

Tabular Data

- **TabNet:** A general model which performs well on a wide range of classification and regression tasks.
- **AutoML Tabular Workflow:** A complete AutoML pipeline for classification and regression tasks.