WISL_No97 978-981-17970-7-1

Naige_Care_Presentation

Based on the provided image, **Option 2 for assignment** outlines three choices:

- **Introduce a new technique** (such as AI agent, Data Agent etc.)
- **Or discuss the ethics issues of a GenAI technique**, such as ChatGPT, VLLM, RAG, AI agent, Embodied AI, GenAI in general, or Claude Code
- **Or a data science Success Stories & Cautionary Tales** (especially new case in LLM era)

The image also notes that a large language model (LLM) is an AI system that processes, manipulates, and generates text, with training involving inputs of large amounts of data (corpus) through the following stages:

- **Data Gathering**: Sources of misinformation, inherent bias with the sources.
- **Data Cleaning and Pre-processing**: Stringent filter (criteria) applied.
- **Model Architecture**: Incompatibility of data with model results in wrong interpretation of cause-and-effect.
- **Pre-training**: Inherent bias with training dataset.
- **Fine-tuning**: Incorrect assignment or adjustment of weights to amplify or nullify to compensate.
- **Evaluation & Deployment**: Misuse and abuse of the output data.

A	B	C	D
Group ID	Presentation time slo	Name/email 1	Name/email 2
	Saturday 8 August		
1	09:00	Ng Kim Ee (Ngkimee@gmail.com)	
2	09:10		
3	09:20		
4	09:30	Andrew Toh / andrew.toh78@gmail.com (Opt 1)	
5	09:40		
6	09:50		
break	break		
7	10:10	Jeremy / jeremysv78@gmail.com	Geraldine Toh / geraldine.to
8	10:20		
9	10:30		
10	10:40	Edwin Koh Wui Kiat, kohpapa@outlook.com on Option 2 [Eldercare]	

Ethics, Bias, and Data Governance in Generative & Agentic AI

Evaluating Data Science Pipelines and Safety Risks in High-Acuity Healthcare (NAIGE Care Framework)

Option 2 Assignment Case Study

Introduction — The NAI 4.0 Governance Framework

Sovereign, non-agentic engineering standards for residential high-acuity environments

Privacy-by-Physics (P-Layer)

Eliminates raw biometrics at the silicon level using 905nm LiDAR voxelization. Captures pure Center-of-Mass (CoM) kinematic vectors without cameras or microphones.

Deterministic Deliberation (N-Layer)

Operates on a 1 kHz master clock enforcing the Sacred Pause™ (25–1000 ms hardware timing) and Tripartite.1x Key™ for high-risk authorization.

Finality & Fail-Safe (D-Layer)

Equipped with the U7 Sovereign Brake™ (severing LDO power rails in under 1.05 ms upon anomaly detection) and the WD117 Enhanced WORM Ledger recording events via OTP silicon fuses and SHA-3-512 hashing.

Data Science Life Cycle & Ethical Vulnerabilities

Systemic risks across LLMs and generative data pipelines

- **Data Gathering Vulnerabilities:** Ingestion of massive text or sensor corpora introduces systemic misinformation and inherent source biases.
- **Data Cleaning & Pre-processing Pitfalls:** Over-filtering or flawed cleaning criteria mask critical edge cases, disproportionately harming vulnerable demographics.
- **Model Architecture & Pre-training Bias:** Incompatibility between data distributions and model assumptions leads to flawed cause-and-effect interpretations.
- **Fine-Tuning & Evaluation Risks:** Misguided weight adjustments or unconstrained autonomous agents risk hallucinations and operational misuse in safety-critical domains.

The Cautionary Tale — Autonomous Agents vs. High-Acuity Risk

Why probabilistic systems fail in vulnerable residential care environments

The Probabilistic Problem

Traditional LLMs and autonomous AI agents operate on probability distributions, making them fundamentally prone to unexplainable outputs and critical hallucinations.

High-Acuity & Privacy Stakes

Managing conditions like dementia, stroke, and diabetes leaves zero room for error. Furthermore, cloud-based audio/video IoT creates severe surveillance and data breach liabilities.

Data Science Solution — Privacy-by-Physics & Non-Agentic Design

Redefining the data pipeline through the 3ZEROS™ Protocol

The 3ZEROS™ Protocol

Zero Camera, Zero Audio, Zero Cloud transmission. Eliminates raw biometric collection at the hardware silicon layer entirely.

Agnostic Data Gathering & Dignity

Utilizes 905nm LiDAR voxelization to capture pure Center-of-Mass (CoM) kinematic vectors. Protects user dignity by processing structural geometry rather than invasive personal identifiers.

Deterministic Deliberation & Algorithmic Guardrails

Replacing open-ended agentic flexibility with hardwired hardware controls

The Sacred Pause™

Enforces a mandatory 25–1000 ms hardware timing delay to evaluate and validate high-risk model outputs before execution.

Tripartite.1x Key™

Requires multi-factor verification (iris scan, cryptographic token, and kinetic foot pedal) ensuring absolute human sovereign oversight.

Fail-Safe Engineering & Regulatory Alignment

Rigorous hardware audit trails and regulatory compliance standards

System & Data Integrity

WD117 WORM Ledger: Records operational events via OTP silicon fuses and SHA-3-512 hashing.

U7 Sovereign Brake: Severs LDO power rails in under 1.05 ms upon anomaly detection.

Regulatory Standards Mapping

Aligns with HSA Class B through D tiers and international standards (ISO 14971, IEC 62304) by embedding governance directly into silicon hardware rather than soft software policies.