1.0 Azoma

1.1 Optimizing Product and Brand Visibility in AI-Driven Search Engines

- Core Challenge: As consumers shift to AI-driven conversational platforms like ChatGPT, Google Gemini, Perplexity, and Amazon Rufus, brands face a new challenge in ensuring their products are recommended and ranked favorably.
- Primary Objective: To design and validate methodologies that systematically improve the visibility and ranking of e-commerce products within these new AI search environments.
- Data Collection: The project involves building a data pipeline to interrogate AI search platforms with product queries, effectively creating an AI-search equivalent of traditional Search Engine Results Page (SERP) data for analysis.
- Methodology: Interns will adapt findings from the academic paper "Manipulating Large Language Models to Increase Product Visibility" to develop techniques for optimizing product descriptions, using models like GPT and Llama 3 for testing.
- Key Deliverable: A prototype solution that automates the generation of optimized content, integrates into e-commerce workflows, and demonstrates a measurable improvement in product visibility.

2.0 CPP Investments

2.1 Application of Financial Machine Learning in Forecasting Company Fundamentals

- Main Goal: To explore how financial machine learning (FML) methods can improve the accuracy and timeliness of forecasting a company's financial performance.
- Data Integration: The project focuses on integrating traditional financial data (e.g., financial statements) with alternative data sources like consumer transactions, app usage, and web traffic.
- Technical Approach: Interns will implement advanced techniques, including feature engineering for panel-to-population prediction, and test machine learning models such as Gradient Boosting and Transformer-based architectures.
- Benchmarking: A key part of the project is to compare the performance of the new ML models against existing baseline statistical models used within CPP Investments' framework.
- Expected Outcome: A working prototype of an FML forecasting model and a quantitative analysis demonstrating the added value of using alternative data in investment decision-making.

3.0 Cresta

3.1 Agent Workflow Automation

- Core Problem: Contact center agents are slowed down by a heavy load of manual, repetitive tasks across multiple systems, leading to errors, inconsistent compliance, and slower customer issue resolution.
- Primary Objective: To build a reusable framework that can automate Standard Operating Procedures (SOPs) end-to-end, aiming for a ≥20–30% reduction in manual steps for agents.
- Methodology: The project involves surveying and comparing different automation approaches, including API integrations, browser automation (e.g., Playwright, Puppeteer), and the use of multimodal LLMs.
- Key Deliverable: A lightweight, reusable workflow layer with guardrails and observability, demonstrated through automating at least three representative SOPs like after-call work or KYC verification.
- Business Impact: Success means improving agent productivity and compliance, while lowering average handling time and operational costs for contact centers.

3.2 Charting and Counting in AI Analyst

- Core Challenge: Generating custom charts and reports from contact center data is a time-consuming, labor-intensive process that requires specialized skills, hindering ad-hoc data analysis.
- Project Goal: To develop an AI-driven capability that allows users to generate analytical summaries and data visualizations (charts, reports) simply by using natural language conversational inputs.
- Technical Approach: The project explores converting ambiguous natural language intent into a structured query plan, using methods like text-to-SQL or LLM-powered tools.
- Key Objective: To design, prototype, and compare methods for both data querying and chart generation, ultimately integrating an ALPHA version of the feature into Cresta's production systems.
- Innovation Focus: The project lies at the intersection of several active research areas, including Agentic AI, Multi-modality, Reinforcement Learning (RL), and post-training, to tackle the novel problem of conversation-driven charting.

6.0 Samsung Research

6.1 Computational Photography & On-Device AI for Mobile Cameras

- Primary Goal: To improve smartphone camera image quality, particularly in challenging scenarios like low-light conditions, nighttime photography, and digital zoom.
- Technical Focus: The project applies state-of-the-art deep learning techniques, such as vision transformers, diffusion models, flow-matching models, and multi-modal LLMs, to solve difficult image restoration and enhancement tasks.
- Core Challenge: A key hurdle is deploying these complex and computationally intensive AI models onto mobile phones, which have limited memory and processing power.
- Hardware Context: The research is closely tied to the performance of the in-camera Image Signal Processor (ISP), with the team working on everything from software-based ISP emulation to training deep learning components for it.
- Real-World Impact: This internship provides an opportunity to contribute to improving the camera technology that will be used by millions of Samsung device users worldwide.

7.0 Sanofi

7.1 Digital Biomarkers

- Project Goal: To develop scalable AI solutions for analyzing multimodal data (images, sensors, omics) to accelerate pharmaceutical research and development.
- Data Sources: The project focuses on "Digital biomarkers" (DBMs), which are collected from wearables, smartphones, and other sensors to provide real-time physiological and behavioral metrics.
- Key Challenge: Handling vast volumes of high-resolution data from multiple sources and formats, such as videos for gait analysis, sensor recordings, and spatial transcriptomics images.
- Methodology: Interns will work with computer vision models for image analysis, time-series analysis for sensor data, and signal processing techniques to extract meaningful insights.
- Application Areas: The developed AI tools will be applied to enhance clinical trials through more accurate and frequent measurements, with examples like motion tracking from videos and scalable analysis of spatial transcriptomics data.

7.2 Bridging Clinical Language Models to Patient Matching for Clinical Trials

- Core Problem: Matching patients from real-world data (RWD) to clinical trials is a major bottleneck in research because trial eligibility criteria are written in a different format and vocabulary than patient records.
- Primary Objective: To prototype an end-to-end pipeline that uses Sanofi's internal Clinical Language Models (CLMs) to automate the process of matching patients to clinical trials.
- Proposed Method: The project involves developing methods to translate unstructured inclusion/exclusion criteria from trial protocols into structured queries or dense embeddings compatible with RWD.
- Evaluation Strategy: The intern will investigate and compare two main approaches: rule-to-representation translation and patient similarity scoring based on embedding vectors.
- Key Deliverable: A validated prototype pipeline accompanied by a systematic framework for presenting patient-matching results, with the potential for the work to be published.

7.3 Sequence and Structural Modeling for Therapeutic Discovery

- Overarching Goal: To leverage machine learning to build advanced computational tools that reduce costs and accelerate all stages of the drug development pipeline.
- Impact Opportunity: Contribute directly to the design of next-generation mRNA vaccines by generating and optimizing novel sequences.
- Technical Approach: Employ a range of methods, from established architectures like transformers and Equivariant Graph Neural Networks (EGNNs) to exploratory techniques like Reinforcement Learning (RL)-guided generation.
- Novel Challenges: Develop bespoke computational methods for de novo enzyme generation and create AI-based tools to replace physics-based modeling of pharmaceutical reaction chambers.
- Collaborative Environment: Interns will contribute as scientists, with the opportunity to publish their findings in top conferences, following the precedent set by previous interns in the group.

7.4 Using Agents on Real World Data for Medical Strategy

- Core Problem: A critical limitation in healthcare data analysis is the inability to leverage Real-World Data (RWD) like electronic health records (EHR) and claims data for real-time, on-demand computation.
- Project Aim: To design and implement an AI agent capable of accessing, interpreting, and computing values from RWD in real time in response to natural language queries.
- Technical Context: This agent will be integrated into an existing multi-agent framework, expanding its capabilities beyond its current scope and enabling it to query a new class of valuable datasets.
- Methodology: The intern will research the latest advancements in AI agents and LLM-powered SQL generation, create a proof-of-concept, and develop a plan for scaling the solution to be production-ready.
- Business Impact: This work will provide medical teams with powerful new tools for data-driven decision-making, ultimately leading to greater patient impact.

7.5 Knowledge Engineering for AI Agents and Enterprise Intelligence

- Core Challenge: As AI agents become more integrated into company operations, there is a pressing need to power them with a consistent, accurate, and centrally-managed source of organizational knowledge.
- Primary Objective: To design and build a robust, centralized knowledge repository that serves as the "brain" for Sanofi's network of AI agents.
- Key Technologies: The project will use Snowflake for advanced semantic modeling and a graph database (e.g., Neo4j) to map and represent complex relationships between different knowledge domains.
- Methodology: The work involves a full-cycle development process: conducting a knowledge audit, designing the system architecture, implementing ETL pipelines, developing APIs for agent access, and creating analytics tools to measure performance.
- Expected Outcome: A comprehensive knowledge engineering system that ensures AI agents across the enterprise provide accurate, relevant, and timely information, improving efficiency and decision-making.

7.6 Application of AI/ML Methods for Target Discovery and Planning Optimization

- Core Problem: Current analytical methods struggle to extract meaningful biological insights from complex, high-dimensional human data, hindering drug discovery and clinical trial design.
- Application Area 1 (Disease Endotypes): Using neural networks (GNNs, VAEs) and unsupervised methods to uncover distinct biological mechanisms and patient subgroups from time-series clinical transcriptomics data.
- Application Area 2 (Target Identification): Leveraging multi-modal data (gene expression, biomarkers, clinical variables) and integrative analysis methods to discover new, actionable therapeutic targets.
- Application Area 3 (Trial Optimization): Collecting internal and external data to develop models for task optimization and risk prediction in clinical trial planning, using AI and probabilistic methods like Bayesian analysis.
- Innovation Focus: The project aims to build upon previous internal research by applying novel computational approaches, such as diffusion-based denoising models inspired by AlphaFold, to push the boundaries of drug development.

8.0 Shopify

8.1 Agentic Memory Graph

- Core Problem: LLM-based agents, like Shopify's Sidekick assistant, often struggle with context and recall over long, multi-turn conversations, as standard vector search fails to capture conversational structure.
- Proposed Solution: To design and implement an "Agentic Memory Graph"—a dynamic graph structure that represents conversational events, entities, and their relationships.
- Technical Design: The graph will use nodes for conversation turns and entities, and typed edges to represent relationships like correction, preference, dependency, and temporal order.
- Primary Objective: To augment the Sidekick AI assistant with a more robust memory system that enables precise, low-latency recall of the most contextually relevant facts.
- Evaluation Method: The new graph-based memory system will be quantitatively benchmarked against standard RAG/vector search baselines using AI judges on simulated conversations.

8.2 Enhancing Deep Research

- Strategic Shift: To evolve the Sidekick AI assistant from being purely reactive (answering direct questions) to being proactive, acting as a trusted business advisor.
- Core Function: The project focuses on an offline research pipeline that proactively analyzes a merchant's shop data to identify hidden opportunities they might have missed, such as sales trends or cart abandonment issues.
- Key Technical Initiatives: To enhance the pipeline's reliability and accuracy through content grounding (to reduce hallucinations), robust error handling, and building an intelligent ranking system.
- Proactive Assistance: The system will generate specific, actionable messages for merchants based on its findings, surfacing opportunities they didn't know existed.
- Methodology: The intern will use a combination of prompt engineering, data analysis of failure logs, A/B testing for the ranking system, and a simulation framework to test action effectiveness.

8.3 Diversifying Admin Search Results

- Current Limitation: Shopify Admin's current search system is based on exact-text lexical matching, which means it cannot handle typos, understand natural language, or provide context-aware results.
- Primary Objective: To design and implement a hybrid search architecture that combines the speed of lexical search with the intelligence of semantic search, all while maintaining sub-100ms latency.
- Key Features: The new system will support fuzzy matching for typo tolerance, semantic search over FAQ content, context-aware reranking of results, and natural language query understanding.
- Methodology: A phased approach will be used, starting with implementing fuzzy lexical search (BM25) and a vector index for semantic search, followed by developing a lightweight reranking model and fine-tuning small language models.
- Business Impact: This project will transform the search experience for millions of merchants, making it faster and more intuitive to find information and manage their stores.

8.4 Enhancing Shopify's AI Assistant with Advanced Code Generation

- Technical Gap: Shopify's Sidekick assistant is effective for merchant-focused tasks but lacks the sophisticated, domain-specific code generation capabilities required for advanced development.
- Three Enhancement Domains: The project aims to significantly improve Sidekick's abilities in: 1) Flow Automation, 2) Theme Development (Liquid, CSS, JS), and 3) App Creation (React/GraphQL).
- Core Methodologies: The approach will involve domain-specific fine-tuning of LLMs on Shopify's codebase, developing custom reasoning models for code planning, and using Retrieval-Augmented Generation (RAG).
- Key Deliverable: A prototype of Sidekick extensions that demonstrate measurable improvements in AI-assisted software development on the Shopify platform.
- Ultimate Goal: To transform Sidekick into an advanced development platform accessible to millions of merchants, regardless of their technical expertise.

8.5 Enhancing Evaluation System for Sidekick

- Core Challenge: As the Sidekick AI agent evolves, its evaluation systems must scale to handle manual prompt engineering bottlenecks, identify hidden feature gaps in user conversations, and manage an expanding library of tools.
- Objective 1 (Prompt Optimization): Integrate the DSPy framework to automatically generate and refine system instructions from conversation examples, reducing manual effort and speeding up iteration.
- Objective 2 (Conversational Intelligence): Build an analytics pipeline to mine conversation logs, using clustering and error classification to surface recurring user issues and unmet needs for product teams.
- Objective 3 (Tool Quality Management): Systematically evaluate the agent's tool library to detect ambiguity and overlap, which can cause agent confusion, and generate recommendations for improvement.
- Overall Impact: To create a comprehensive evaluation system that improves agent quality, accelerates development velocity, and enables data-driven product decisions for Sidekick.

8.6 Improving Assistant Simulator for E2E Production Evaluation

- Evaluation System: Shopify uses an internal system called MerchantLLM that tests proposed code changes by replaying real production conversations against a release candidate.
- Core Challenge: The system faces a complex trade-off: it must know when to faithfully adhere to the original conversation path and when to intelligently deviate from it if a code change materially alters the assistant's response.
- Added Complexity: The simulator must also handle changes in data resources (e.g., a user's profile or product inventory) that may have occurred between the original conversation and the replay.
- Project Goal: To improve the MerchantLLM simulator to make it a more reliable and efficient end-to-end evaluation system for all of Sidekick's capabilities.
- Methodology: The intern will develop semantic similarity metrics to detect functionally different responses, build a meta-evaluation framework to test the simulator itself, and shift the evaluation paradigm to better isolate the impact of code changes.

8.7 AI-Powered Personalization for Shopify Admin Home Feed

- Current State: The Shopify admin home feed is static, showing the same content to all merchants regardless of their unique business situation (e.g., stockouts, traffic decline).
- Future Vision: To transform the home page from a static dashboard into an intelligent, personalized command center that proactively surfaces the most critical information and actions for each merchant.
- Technical Solution: The project involves building and deploying an AI-powered ranker that personalizes the feed's content based on real-time shop context and merchant priorities.
- Key Signals: The ranker will incorporate various shop signals like inventory levels, traffic trends, order velocity, and seasonal patterns to determine what content is most relevant at any given moment.
- Performance Constraint: A critical requirement is to maintain an instant user experience, with page load times remaining under 200ms, which necessitates efficient feature engineering and low-latency model inference.

9.0 Unilever

9.1 A Framework for Auditable Strategic Reasoning via Knowledge Graph-Augmented LLMs

- Core Problem: Standard Retrieval-Augmented Generation (RAG) systems fail at the complex, multi-hop reasoning required for high-stakes business strategy and lack the step-by-step explainability needed for verification.
- Primary Objective: To research and build a novel reasoning system that functions as a verifiable "junior AI strategist," capable of synthesizing insights from vast, disparate data sources.
- Technical Architecture: The proposed solution combines a unified Knowledge Graph, a high-performance Graph-Native Retrieval (GNR) engine, and an auditable reasoning framework.
- Novel Methodology: The project's core innovation is an iterative "Graph Chain-of-Thought" (Graph-CoT) agent whose reasoning is validated at each step by a KG-RAR-inspired verifier module to ensure auditability.
- Key Outcome: An end-to-end system that delivers trustworthy, transparent, and auditable insights for complex strategic inquiries, moving beyond the limitations of opaque RAG systems.

9.2 A Framework for Empirical Benchmarking of Agentic AI Orchestration Patterns

- Research Gap: While agentic AI is a promising field, there is a critical lack of empirical understanding of how different multi-agent architectures (e.g., hierarchical, sequential) perform on real-world tasks.
- Primary Goal: To research and build a standardized benchmark framework for rigorously evaluating the performance trade-offs of various agent orchestration patterns.
- Key Metrics: The framework will measure and compare architectures based on critical business metrics: latency (speed), solution quality, and operational cost (API tokens).
- Methodology: The intern will develop a 'test harness' of standardized business tasks and a set of modular 'utility agents' that can be plugged into different orchestration systems for comparison.
- Key Deliverable: A comprehensive benchmark report that provides a data-driven "playbook" for choosing the most effective agentic architecture for a given business problem.

9.3 Next-Generation Material Price Forecasting

- Business Need: Accurate forecasting of raw material prices is critical for Unilever's financial planning and strategic sourcing, but conventional models are reaching their predictive limits.
- Primary Objective: To architect a state-of-the-art forecasting engine that significantly outperforms existing benchmarks by pioneering the application of novel AI techniques.
- Innovation 1 (Causal Inference): The project will use causal inference techniques to precisely model the impact of major market shocks, rather than treating them as simple outliers.
- Innovation 2 (Foundation Models): A key focus is to be among the first to deploy a fine-tuned Time Series Foundation Model (like Google's TimesFM) for an industrial application, leveraging few-shot learning for superior performance.
- Success Criteria: A production-ready prototype that demonstrates a quantifiable lift in forecast accuracy and delivers reliable probabilistic forecasts (prediction intervals) for sophisticated risk analysis.

10.0 Vanguard

10.1 Agentic AI for Adaptive Fraud Defense

- Core Problem: Financial fraud is growing more sophisticated and automated, and traditional detection systems with static rules struggle to adapt to rapidly evolving attack strategies.
- Strategic Vision: To move beyond point solutions toward an adaptive, end-to-end defense lifecycle that continuously learns, anticipates vulnerabilities, and improves its response over time.
- Project Goal: To design and evaluate an agentic AI framework that integrates detection, simulation, and investigative reasoning across the entire fraud defense process.
- Methodology: The framework will feature both defensive agents (for detection and reasoning) and offensive agents (for simulating novel fraud behaviors to stress-test the system), creating an iterative feedback loop.
- Key Outcome: A prototype of a self-evolving fraud defense ecosystem that enhances the resilience, adaptability, and explainability of modern fraud prevention.

10.2 Adversarial Attacks on Financial AI Agents

- Compliance Risk: Client-facing financial AI agents (e.g., chatbots) must adhere to strict regulations, but subtle user inputs can cause them to provide inappropriate advice, creating significant legal and reputational risk.
- Primary Objective: To develop a systematic framework for generating and evaluating "regulatory-adversarial" prompts specifically designed to test the compliance boundaries of financial AI agents.
- Methodology: The project involves converting financial regulations into machine-readable rules, algorithmically producing prompts that challenge the AI's compliance, and establishing metrics to benchmark model performance.
- Novelty: This work extends traditional adversarial testing into the highly regulated financial context, algorithmically integrating formal constraints into the attack generation process.
- Key Deliverable: A "Financial Regulatory Jailbreak Benchmark" consisting of an attack algorithm, a prompt dataset, and an evaluation framework to help institutions improve the safety and robustness of their AI systems.