Understanding Meta's Llama2- Quadwave Perspective

Llama2, an **Open source LLM** by Meta, is capable of providing optimized NLP focused solutions for R&D & Commercial use cases.

- Ranging from 7B to 70B parameters
- Optimized for Local running
- Safety red teaming
- Open innovation and collaboration
- Improved Contextual Understanding
- Benchmarks outperformance

What can Llama2 do?

- Natural Language Understanding and Generation of human like text
- Market insights generation
- Creative content generation and personalization
- Research and Academia support for individuals and institutions
- Sentiment analysis of text-based inputs for further insights generation
- Coding and reasoning abilities

What can Quadwave do in Llama2:

Quadwave's Llama2 competencies are primarily targeted providing enhanced efficient & innovative solutions to customers across various industries. These services are bucketed into:

Llama2 Product Evangelization

- Workshops, Seminars & Webinars
- Model Comparative Studies
- Training & Readiness Sessions
- Community Networking & Initiatives
- Product Feedback & Supports

Technical Consultancy

- Development & Integration
- Implementation, Fine-tuning Services
- Capacity Planning & Optimization
- Model Evaluations & Extensibility
- Support & Hypercare

Professional CoE Services

- Use Case Exploration & Mapping
- PoC on real-life scenarios
- Troubleshooting & Issue Resolutions
- Guidance & Best practices
- Develop Position Papers, etc.

Integrated LLM/GAI Solutions

- Conversational Solutions i.e. Chatbots
- Content Generation & Summarization
- Knowledge Management Solutions
- Industry-specific NLP Solutions

Anti-scamster LLM Solution- “athenaGUARD”

Objectives: “athenaGUARD” should be able to guide & Help citizens to enjoy digital financial journey by guarding against scamsters- by providing to-the-point anti-scam information & helping to address scam-concerns.

Target Audience: Citizens intending to access digital financial apps but are afraid of ever-increasing scam/s. We have categorized their personas:

- Awareness Seekers
  (First-time digital financial app users)
- Affected Users
  (Affected by Scams & looking for quick help)
- Security Providers
  (Banks, Government Agencies Security Team)

User Requirements:

- Informative Resources
  (Best Practices, Guidelines, Do & Don’ts, etc.
- Guided Interaction
  (Bank Helplines, Forms, Consumer Helpers, etc.)
- Policy & Compliance Standards
  (Policies, Anti-scam steps, Regulations, etc.)

Solution: “athenaGUARD” is an user-centric LLM solution that uses Llama2 model to extract scam-specific information and fine-tune localized policy knowledge base to provide real-time need-specific solutions to the users. Given below solution differentiators:

- User-tailored Functionalities
  (FAQ, Q&A, Raising Scam Queries, etc.)
- Smart Defaults for Parameters
  (Precision, Diversity, Sensitivity)
- Capture patterns in policy writing
  (Policy Guidelines Identified)
- Optimized Resource Utilization
  (Model optimization based on resource availability)
- User query soft prompting
  (User-friendly Natural Query auto-enhancement)
- Context aware insights
  (Coherent and responsive to context)
- Precise & Factual Information Retrieval
  (Accurate and consistent context and output)
- User centric approach
  (Self adaptation based on human feedback)
- Up to date policies and compliance knowledge
  (Real time adaptability)

Solution Overview- “athenaGUARD” use cases

Awareness Seekers

Background Behavioral traits:

- Brian is a 45-year-old government employee responsible for IT security awareness programs.
- He is tasked with educating colleagues and the public about the risks associated with spam and cyber threats.
- Proficient in IT, capable of understanding and implementing advanced security measures.

Goals and Needs:

- Requires access to educational resources to enhance his knowledge of emerging scam threats.
- Wants to integrate the real-time support system with the government's existing awareness programs.
- Seeks regular updates on scam trends to tailor awareness campaigns effectively.

Affected Users

Background Behavioral traits:

- Ava is a 32-year-old professional who relies heavily on her mobile phone for work and personal communication.
- She has been receiving an increasing number of scam calls and messages, leading to frustration and concern about her privacy.
- Comfortable with basic smartphone operations but may need guidance on advanced settings.

Goals and Needs:

- Wants immediate assistance in dealing with scam issues.
- Wants step by step information on how to file a complaint to how to receive a solution.
- Expects informative interactions to stay updated on the latest scam-related threats.

Security Providers

Background Behavioral traits:

- Olivia is a 38-year-old security manager working for a financial institution.
- She is responsible for implementing and maintaining security measures to protect customers from scam-related threats.
- Highly proficient in IT security and expects a system that aligns with industry standards.

Goals and Needs:

- Requires integration capabilities with the internal policies of the financial institution.
- Expects a comprehensive real-time support system to promptly address and mitigate scam-related issues by generating
- Seeks informative interactions to stay ahead of the evolving nature of scam threats.

Measurement index:

- User Issue Resolution rate
  (Resolution confirmation follow-up, Escalation rates)
- Integration effectiveness rate
  (Assess by verifying sources retrieved)
- Consistency in Policy Adherence
  (Adherence Uniformity Assessment)
- Completion Rate evaluation
  (step-by-step success metric)
- Chat response time
  (Speed of Assistance Evaluation)
- Context aware insights
  (Query response relevance)
- Feedback integration accuracy
  (Validating learning/adaptation from Human feedback)
- User engagement rate
  (Participation and Interaction Measurement
- Functionality switching efficiency
  (delay time when adapting functionalities)

Thank You