

## Why Meta/Why This Role

Meta's open-source Llama models and commitment to building robust, accessible AI infrastructure make it the most compelling environment for deploying next-generation generative AI. The company's \$14.3B investment in Scale AI, the recruitment of Alexandr Wang, Nat Friedman, and Daniel Gross, and the hiring of eight leading Chinese AI researchers show a clear focus on research leadership and global product impact. My experience aligns with Meta's mission to democratize AI and build scalable, trustworthy systems for worldwide use.

**Top 3 Bullet Points:** – Processed 2M+ km<sup>2</sup> satellite imagery daily for SpaceKnow, serving 350,000+ Bloomberg terminals. – Fine-tuned Llama 3.1 on thousands of internal documents for SpaceKnow's enterprise chatbot. – Built infrastructure supporting 1,000+ QPS web data acquisition for LLM pretraining.

## GenAI Experience

I have led large-scale generative AI projects from infrastructure to deployment, including fine-tuning Llama 3.1 for specialized enterprise use. My work covers high-throughput data pipelines, distributed ETL, and robust automated preprocessing for multi-modal datasets. I have hands-on experience with domain-specific LLM adaptation, web-scale scraping, and ML framework optimization across cloud and on-prem clusters.

**Top 3 Bullet Points:** – Fine-tuned Llama 3.1 on thousands of internal SpaceKnow documents for an internal chatbot. – Built automated preprocessing pipelines handling 14TB/day on 256-node clusters at The Climate Corporation. – Implemented AutoGPT and BabyAGI frameworks achieving 92% accuracy in privacy risk scoring at Spartacus.

## Product Sense

My product sense is grounded in delivering actionable AI solutions for enterprise and scientific users. I focus on user needs, data quality, and transparent outputs, creating products that turn complex analytics into accessible insights for decision-makers. This approach has driven adoption across finance, agriculture, and defense sectors.

**Top 3 Bullet Points:** – Created SpaceKnow Satellite Activity Index for 350,000+ Bloomberg subscribers. – Reduced corn yield prediction error from  $\pm 15.2\%$  to  $\pm 12.1\%$  across 6.2M acres at The Climate Corporation. – Built privacy systems reducing user data exposure by 83% within 30 days at Spartacus.

## Approach to Design

### 1. Design for **Immediate Business Value**, Not Technical Elegance

Pavel consistently chooses solutions that deliver measurable business outcomes over technically impressive architectures. At SpaceKnow, instead of building a general-purpose satellite analysis platform, he created specific products like the **China Satellite Manufacturing Index** that financial analysts could immediately use for trading decisions. At The Climate Corporation, he focused on the single metric that mattered to farmers: **reducing yield prediction error from  $\pm 15.2\%$  to  $\pm 12.1\%$  across 6.2 million acres**. At Spartacus, rather than building a comprehensive privacy platform, he targeted the specific problem of **data exposure** and delivered **83% reduction within 30 days**. His approach is to identify the one number that matters most to users and optimize everything around improving that number. Technical sophistication is only valuable if it directly improves business outcomes.

### 2. Build Systems That Handle **Real-World Data Messiness**

Pavel's experience shows he designs for the **chaos of production data** from day one. His satellite imagery platform had to **"combine imagery from multiple providers despite imagery being in very different formats"** - a classic real-world problem where data never arrives clean or standardized. At NASA, processing **1.7 million star light curves** required handling instrument failures, cosmic ray hits, and calibration drift. His scraping infrastructure uses **RedisBloom for deduplication** and **proxy rotation** because web data is inherently unreliable and hostile to automated collection. Rather than assuming clean inputs, he builds preprocessing, validation, and error handling as core system components. His systems work with messy, incomplete, and contradictory data because that's what exists in the real world.

### 3. **Optimize for Human Decision-Making Speed**

Pavel designs interfaces and outputs that minimize the time between data analysis and human action. Bloomberg Terminal users needed economic indicators they could trade on immediately, not raw satellite imagery to interpret. His systems delivered **sub-second query response times** while processing **2 million square kilometers daily** because financial decisions happen in minutes, not hours. At Spartacus, the privacy risk scoring had to analyze **250+ data points per user** and present clear recommendations, not technical reports. His NASA work improved **photometry precision from 42ppm to 29ppm** specifically to enable faster exoplanet candidate identification by astronomers. The pattern: he optimizes for the speed of human decision-making, not system throughput. Users need answers they can act on immediately, not comprehensive analysis they need to interpret.

## Stakeholder Management

I have managed partnerships with global financial, energy, and defense leaders, delivering high-value AI products and coordinating cross-functional teams. My experience includes working with demanding clients in sectors that require the highest standards of reliability, security, and precision. – Finance: BlackRock, Bloomberg, JPMorgan, GoldmanSachs – Energy: Exxon, Trafigura – Defense US DOD and NATO

**Top 3 Bullet Points:** – Managed partnerships with BlackRock, JPMorgan, Goldman Sachs, Exxon, Trafigura, and multiple Departments of Defense. – Built enterprise relationships generating \$10M+ ARR and serving 350,000+ Bloomberg subscribers. – Led contracts with NATO and ESA for classified defense intelligence projects.

## Personal Strengths

My technical strengths include distributed systems, ML pipeline optimization, and large-scale data engineering. I excel at translating technical advances into business value and building teams that deliver complex AI products on schedule and budget.

**Top 3 Bullet Points:** – Built infrastructure processing 14TB/day on 256-node clusters with automated scaling and fault tolerance. – Achieved 92% accuracy in privacy risk scoring while analyzing 250+ data points per user in real-time. – Managed cross-functional teams of 50+ engineers and data scientists delivering complex AI products.