Anjan Goswami

San Francisco Bay Area, CA
Objective: *AI, Search, and Data Executive*

goswami.anjan@gmail.com
https://www.linkedin.com/in/goswamianjan/
https://www.smartinfer.com/

Executive Summary

- **AI & Engineering Executive:** 20+ years building 0-to-1 AI products across enterprise and consumer platforms. Led large-scale initiatives in generative AI, NLP, computer vision, search, and recommender systems.

- **Production Delivery at Scale:** Built and led cross-functional teams delivering production AI systems used by tens of millions of users, driving measurable business outcomes and sustained product advantage.

- **Systems-first Builder:** Deep expertise translating research into scalable, high-performance AI platforms under real-world constraints (latency, cost, quality, safety, compliance).

Key Achievements

- **LLM-Based AI for Microsoft PowerPoint (2024–Present):** Architected and delivered agentic LLM systems for slide generation, editing, evaluation, and on-brand grounded content. Built image-based slide generation and template selection pipelines. Founded and scaled the AI & Data team (1→16) in one year; delivered Copilot capabilities driving **20%+ user engagement** and informed Office-wide investment decisions through strategy whitepapers and benchmark-driven model evaluations.

- **Data-Centric Document Intelligence at Adobe (2019–2023):** Designed and deployed a new high-performance PDF semantic parsing system to replace a legacy object-detection + rules pipeline.
    - Achieved **2× latency improvement** and processed **10 documents end-to-end in ¡10 seconds**.
    - Improved accuracy from **45% → 65%**, matching transformer-based models (LayoutLM at **63%**) with significantly lower cost and complexity.
    - Built a **privacy-preserving active learning system** using short-lived production data, dark launches, automated evaluation, and targeted human labeling—fully GDPR compliant.

- **ML Platforms at Salesforce (2016–2019):** Built the Community Cloud ML organization from scratch. Delivered feed ranking, multiple recommender systems, spam detection, Q&A, chatbots, and enterprise search—driving **20%+ engagement lift** and accelerating adoption of Transformer-based models.

- **Commerce Search at Walmart (2014–2015):** Led an 80+ person team to rebuild the e-commerce search stack from scratch, including ranking, query understanding, spell correction, and type-ahead—resulting in **23% revenue growth** and **17% sales uplift**.

- **Recommender Systems at UpWork (2013–2014):** Built a near real-time contractor recommender system from scratch (**+4% hiring rate**); established experimentation and analytics platforms supporting M&A diligence.

- **Search at eBay & Amazon (Earlier Career):** Architected ML ranking for eBay Cassini (**+6% revenue**) and image search for fashion; at Amazon A9, developed global ranking infrastructure and relevance evaluation systems.

Areas of Expertise

- **AI & Information Retrieval:** Built and shipped large-scale search, ranking, recommendation, and multimodal AI systems across enterprise and consumer products.

- **AI Systems & Execution:** End-to-end AI platforms under real-world constraints—latency, throughput, cost, evaluation, safety, and compliance.

- **Leadership & Responsible AI:** Building high-performing AI organizations, advancing responsible AI practices, and aligning applied science with strategic product outcomes.

Professional Experience

Head of AI and Data $\to$ Senior Principal Scientist, Microsoft PowerPoint (Office 365) *Mar 2024 – Present*

- **AI Strategy & Agentic Architecture:** Defined and executed the end-to-end AI and data strategy for PowerPoint Copilot. Architected LLM-based agentic systems enabling slide generation, editing, rewrite, summarization, Q&A, reference-grounded and on-brand content generation, and structured Office XML creation.

- **Business Impact:** Delivered Copilot capabilities driving **20%+ user engagement** and measurable productivity gains across global enterprise and prosumer users.

- **Org Formation & Talent Bar:** Built and scaled an applied science and data organization from **1 to 16** within one year. Recruited senior AI talent with strong research records or deep production experience; established a hiring bar focused on conceptual depth, critical thinking, and rigorous problem solving.

- **Evaluation, Grounding & Quality Systems:** Designed end-to-end evaluation frameworks for slide generation and editing, covering relevance, factual grounding, brand compliance, layout quality, and regression detection across model, prompt, and orchestration changes.

- **Multimodal Intelligence & Design Systems:** Led development of image-based slide generation, template selection, and design recommendations by integrating LLM reasoning with visual, layout, and template-retrieval signals.

- **Infrastructure, Safety & Model Lifecycle:** Partnered with Microsoft Research on foundation-model alignment, safety evaluation, and deployment readiness. Improved inference latency and cost efficiency through Azure GPU tuning, caching, and retrieval-path optimizations.

- **Strategy & Competitive Intelligence:** Authored internal AI strategy whitepapers and conducted frontier model evaluations and benchmark-driven capability assessments to inform multi-year product and investment decisions across Office.

- **Senior Principal Scientist Scope (Dec 2025–Present):** Expanded mandate to frontier model evaluation and technical direction. Led rigorous assessments of diffusion-based language models versus autoregressive LLMs for PowerPoint NLP workloads, informing a decision to defer adoption due to determinism and schema-control gaps. Rebuilt infographic image-generation evaluation to address score saturation and dataset bias; demonstrated Nano Banana Pro outperforming Imagen variants (**5 pts**) and Flux (**10 pts**) on presentation-layer quality, while identifying unresolved gaps in instruction adherence and structural completeness—directly shaping subsequent model selection and control strategy.

Founder, Stealth Healthcare AI Startup *Sep 2023 – Feb 2024*

- Founded and built a healthcare-focused AI prototype exploring LLM-based patient education and medical literacy.

- Designed end-to-end architecture including retrieval-augmented generation (RAG), domain grounding, and evaluation safeguards for factuality and safety in medical content.

- Led rapid prototyping to validate product and technical feasibility; paused further development to return to large-scale applied AI leadership roles.

Director of Machine Learning & Engineering, Adobe *Jul 2019 – May 2023*

- **Organizational Leadership & Re-Architecture:** Directed a global 40–60 person applied science and engineering organization delivering AI capabilities across Acrobat, Sign, Stock, Creative Cloud, and Experience Cloud. Re-organized Document AI from a centralized functional science model into product-aligned teams (3 product teams + 2 horizontal teams for data and edge models), improving alignment and throughput between research and product engineering.

- **Data-Centric Document AI at Scale:** Led the design and production deployment of a new document-understanding system built to meet strict latency and throughput SLAs. When transformer-based approaches (e.g., Detectron/LayoutLM) met accuracy targets but failed performance constraints, architected a compact-model pipeline combined with a data-centric training loop—achieving **2$\times$ latency reduction**, **3–4$\times$ throughput gains**, and improving model quality from **45% to 65%**, matching transformer-level accuracy in production.

- **Privacy-Preserving Active Learning System:** Designed a GDPR/CCPA-compliant active learning framework using short-lived production data, privacy-preserving embeddings, public PDF analog matching, dark launches, and automated evaluation. High-impact samples were selectively routed for human labeling based on evaluation signals, enabling continuous quality improvement without retaining sensitive customer data. System was deployed across Acrobat and Adobe Sign pipelines.

- **Generative & Retrieval-Based Document Intelligence:** Led development of **Acrobat Assist**, delivering retrieval-augmented summarization and Q&A for PDFs. Built a privacy-centric Azure-based ML pipeline enabling rapid iteration on document intelligence features while meeting enterprise compliance requirements.

- **Multimodal Search & Recommendations:** Architected CLIP-based semantic search for Adobe Stock and ML-driven recommendation systems for Creative Cloud, improving recall, precision, and content diversity at global scale.

- **Data Platforms & Cost Efficiency:** Built a unified audience-discovery platform for Adobe Experience Cloud using fast nearest-neighbor retrieval and scalable clustering. Simplified pipelines and parallelized execution, reducing infrastructure costs by **4×** while improving marketer adoption and pilot customer satisfaction.

Director of Machine Learning & Engineering, Salesforce _Jul 2016 – Jul 2019_

- **0→1 ML Organization Build:** Founded and scaled the first applied ML organization for Salesforce Community Cloud, recruiting and developing the initial modeling teams and establishing end-to-end ownership across modeling, data pipelines, evaluation, and deployment. Grew the organization to 40+ ML scientists and engineers across four geographically distributed teams (SF, Bellevue, and EU).

- **Feeds, Recommendations & Engagement (Community Cloud):** Designed and shipped ML-based feed ranking, multiple recommender systems, and personalization pipelines, driving **20% improvements in engagement and DAU-related metrics** across large enterprise customer deployments.

- **Trust, Safety & Forum Intelligence:** Built large-scale spam and abuse detection systems, forum Q&A retrieval models, and content-quality classifiers operating over millions of users and billions of interactions, improving content health and user trust in community ecosystems.

- **Service Cloud AI Roadmap & New Teams:** Co-defined the Service Cloud AI roadmap in partnership with an acquired startup founder and senior product leadership. Formed and led a new ML team focused on Service Cloud intelligence, delivering article recommendation, virtual assistant capabilities, and AI-driven case handling.

- **Adoption-Focused AI Systems:** Introduced vertical-specific AI packages and instrumentation to determine whether Einstein features were effective for a given customer's data, significantly improving feature adoption and customer satisfaction. Delivered AI-assisted case completion workflows where models auto-filled service case data, reducing agent effort and resolution time.

- **Conversational & NLP Systems:** Led development of enterprise-grade chatbots, dialogue recommendation systems, and early BERT-based NLP models for Service and Community Clouds, accelerating adoption across retail, banking, and hospitality customers.

- **Governance, Privacy & Compliance:** Partnered with legal, security, and ethics teams to establish GDPR- and HIPAA-aligned data governance, model review, and deployment standards for customer-facing AI systems.

- **Enterprise & Product Alignment:** Worked directly with enterprise customers, product leadership, and GTM teams to translate business problems into ML formulations, ensuring solutions delivered measurable customer and business value.

Executive Consultant (via SmartInfer LLC) _Oct 2015 – Jun 2016_

- **Etsy:** Advised senior leadership on search ranking architecture, query understanding, and relevance evaluation; influenced roadmap decisions for ML-based discovery systems.

- **Neurotrack:** Led applied computer vision work improving Alzheimer's diagnostic accuracy by **20%** through enhanced pupil-tracking algorithms and model validation.

Director of Search Science, Walmart E-commerce _Jun 2014 – Oct 2015_

- Led an end-to-end AI transformation of Walmart's commerce search, directing **80+ engineers and scientists** across ranking, recall, and query understanding.
- Re-architected ML-based ranking, demand-aware retrieval, and type-ahead systems, driving **+23% revenue uplift** and **+17% sales conversion**.
- Built the applied science org and talent pipeline, including university partnerships and expansion of the Bangalore R&D center.

Open Source & Applied Research

- **Thinker (Open Source):** Unified LLM inference router with registry-driven model selection, schema-based request language (ThinkerQL), secure credential management, and full observability; supports OpenAI, Anthropic, and Ollama. github.com/smartinfer/thinker
- **LLM Systems R&D:** Early architectural work on **Boson LLM**, a declarative training–inference orchestration framework; explored compact reasoning models for structured tasks.
- **Whitepaper:** *Data Engines for Autonomous Vehicles: A Reference Architecture* — scalable designs for scenario mining, synthetic data generation, multimodal ingestion, and continuous-learning pipelines. smartinfer.com/data-engine-paper
- **NeurIPS 2025 Workshop:** H. Kang, E. Bao, A. Goswami. *VLM-SlideEval: Evaluating VLMs on Structured Comprehension and Perturbation Sensitivity.* arXiv:2510.22045

Earlier Career: Elance–UpWork, eBay, Amazon A9, Microsoft *2005 – 2014*

- **Elance–UpWork (Director):** Built contractor ranking and recommendation systems from scratch, increasing hiring rates by **4%**. Established experimentation and causal inference frameworks (A/B tests, quasi-experiments) to quantify model impact and guide marketplace policy decisions.
- **eBay:** Transitioned core search to ML-based Cassini ranking, delivering **+6% revenue**; architected early fashion image-search pipelines.
- **Amazon A9:** Developed global ranking infrastructure and relevance evaluation systems supporting large-scale e-commerce search.
- Applied economic modeling principles—elasticity analysis, demand forecasting, and incentive-aware optimization—to improve discovery, ranking, and marketplace outcomes across buyer–seller platforms.

Patents, Publications & Thought Leadership

- **Patents (10+):** Search relevance, ranking systems, personalization, and recommendation algorithms. Representative patents include US 1020369, US 20120303615, US 20120221557. Full list: smartinfer.com/patents
- **Publications:** IEEE Big Data (2015), SIGIR ECOM (2018), NeurIPS Workshop (2025). Full list: smartinfer.com/publication/academic_publications.html
- **Essays:** Systems architecture, agentic AI, applied science leadership, evaluation methodologies. smartinfer.substack.com and smartinfer.com/technical_articles

Education

- **Ph.D., Computer Science** — University of California, Davis (AI, Distributed Systems)
- **M.S., Computer Science** — The Ohio State University (Distributed Data Mining)
- **M.Tech, Mechanical Engineering** — IIT Kanpur (Robotics, Vision, Planning)