

Sriram Selvam

SENIOR SOFTWARE ENGINEER - MICROSOFT

+1 (801)-448-5824

| selvamsriram@gmail.com

| srirams.dev

| linkedin/in/sriramselvam

Education

University Of Utah

Salt Lake City, Utah

M.S in Computer Science (GPA 4.0/4.0)

Aug 2018 – May 2020

- Advisor: Prof Robert Ricci
- Thesis: [Switch Assisted Peer To Peer](#)
- Recipient of Graduate Research Assistantship for the entire course duration (Monthly Stipend + Full Tuition Fee Waiver)
- Relevant Courses: Machine Learning, Natural Language Processing, Advanced Algorithms, Data Mining

Government College of Technology

Coimbatore, India

B.E in Computer Science (GPA 7.8/10.0)

Aug 2007 – May 2011

- Within top 2% rank in the state for engineering admissions with a score of 196.25/200
- Relevant Courses: Data Structures and Algorithms, Operating Systems, Digital Image Processing, Computer Networks

Experience

Microsoft Corporation

Redmond, Washington

Senior Software Engineer

Sep 2024 – Current

BING GENERATIVE SEARCH ([Sample](#))

- Designing and building GPT data generation framework for matching the appropriate multimedia and answer cards with LLM generated text summary.
- Work includes fan-out query generation, grounding data collection, data transformation as prompt inputs, LLM result extraction. Pipeline has been used to build >**150,000** result pages.
- Worked with various segment teams at Bing to onboard and match their answer cards in GenSERP.
- Built evaluation metrics for LLM performance on RAG hallucinations, groundedness and reasoning tasks like answer matching (AM Leo).
- Improved the performance of matching as measured by AM Leo metric by more than **30%** through combination of prompt engineering and input massaging and annotations.
- Mentored intern to develop Generative Layouts for GenSERP, defined atomic elements and structured blocks for the UX. Schematized and reformatted data to pass text, media pairs to UX.
- Patent filed with US PTO for Generative Search For Text Documents.

Software Engineer II

Jun 2020 – Aug 2024

BING CREATOR & PERSONAS ([Sample](#))

- Developed Customization Panel UX on SERP, Backend and Dynamic Prompts and Few Shot Learning examples.
- Expanded coverage to new segments like coding, helping users perform tasks like code complexity on generated code.
- Increased engaged DAU by **15x** through expansion of supported segments.
- Designed and implemented a Falcon service API for use across multiple Bing teams, enabling modifications to LLM response qualities such as style, content, and creativity.
- Supported the Companion Answer feature, which co-occurred with news articles to provide enriched perspectives and contextual insights.

GPT INTEGRATION IN BING

- Enhanced Bing's Knowledge Graph, Satori, by leveraging GPT-4 to bridge the gap between strict ontology-based knowledge graph Satori and contextual entity details.
- Developed integrated data pipeline pulling missing contextual details from GPT-4 while respecting ranking priorities and overrides enforced per segment.
- Powered new experiences on entity refinements and AI generated short form videos ([Sample](#)) using the newly added entity contextual details.

ENTITY REFINEMENT ([Sample](#))

- Expanded the coverage of the search tabs feature, displaying relevant tabs like Cast, Plot, Watch, Reviews, Trivia, etc., for specific searches such as movie, tv show names.
- Revamped the UX and data components for the search tabs experience across segments like Movies, TV Shows, Celebrities, Health, and more.
- Developed a custom shopping experience to expose best deals from various retailers for Black Friday 2021.
- Owner for the data generation pipeline for **millions** of actively evolving entities, ensuring high data quality.

Software Engineer Intern

May 2019 – Jul 2019

BING DICTIONARY

- Leveraged BERT embeddings and classical machine learning models to classify web passages as definitions, using a training corpus of dictionary definitions (positive labels) and filtered web crawl text (negative labels).
- Built a stage 2 model to extract the entity being defined from the definition passage.
- Presented findings to FTE team at the end of internship to carry on integration after internship completion.

University of Utah

Salt Lake City, Utah

Graduate Research Assistant

Aug 2018 – May 2020

CLOUDLAB

- Worked with the CloudLab team under Dr. Ricci to identify issues with the OS Image distribution issues within the lab's datacenter environment.
- Proposed and implemented a new network protocol Switch Assisted Peer to Peer to alleviate stress on trunks during large scale image distribution.
- Explored the abilities and limitation of switches that are built on Open Network Operating System and their support for control plane data manipulations and throughput.

IP Infusion

Bangalore, India

Technical Lead Engineer

Jul 2016 – Aug 2018

PCEP & IP FAST REROUTE

- Led a team of 3 members and implemented Path Computation Element Protocol as in RFC 8281 and RFC 8231 for Ciena Corp. The implementation allowed ciena to create and manage MPLS tunnels from SDN Controllers.
- Led and developed a unification logic of IP and MPLS routes in ZebOS, which led to the successful delivery of BGP Labelled Unicast to Ciena Corporation.
- Single-handedly implemented the IP Fast Reroute Feature in IS-IS routing protocol as mentioned in RFC 5286. Successful implementation resulted in zero traffic loss during protocol convergence on network failures.

Senior Software Engineer

Feb 2015 – Jun 2016

MC-LAG & VRF

- Designed and developed a solution for running Rapid Spanning Tree Protocol (RSTP) on Multi-Chassis Link Aggregation protocol (MC-LAG or MLAG). The solution developed helps data center networks to avoid broadcast traffic loops and resultant traffic meltdowns.
- Added new changes in the Linux Kernel 3.16.7 (Debian) to uplift Layer 2 control protocol packets to CPU. This work helped the company to develop a simulation testbed for testing network protocols on VM instead of routers and switches.

- Implemented VRF support for multiple host protocols including SSH, TELNET, NTP, SYSLOG, SFTP, and SCP. The project was developed using SNAT, DNAT, and IP Tables in Debian Linux in a span of 2 weeks. This work facilitated the company to win the product evaluation at Apple Inc.

Software Engineer

Nov 2013 – Jan 2015

TRILL & MC-LAG

- Extensively enhanced and optimized datacenter protocol TRILL and Multi-Chassis Link Aggregation for Fujitsu Japan's DC. The work completed reduced network convergence time on the data center deployment from 500ms to 5ms.
- Spearheaded multiple bugs fixing activities in VLAN, IGMP snooping modules by fixing more than 50 bugs high severity bugs in 2 months of time.

Aricent Group

Chennai, India

Software Engineer

Oct 2011 – Nov 2013

NTP & EPON OLT

- Developed multiple enhancements for line-card provisioning on Aricent Group's Intelligent Switching Solution for Sumitomo Electric, Japan's EPON OLT product FSU 7100 used by Time Warner Cable.
- Implemented NTP support on Aricent's switching solution by utilizing socket communications with NTPD from open NTP.

Skills

- **Programming:** C, C#, Python, Scope Language
- **Libraries:** Pandas, NumPy, Matplotlib, Scikit-Learn, Tensorflow, Keras
- **Data Science:** LLMs, Prompt Engineering, Fine-Tuning, Regressions, Classification, Hyperparameter tuning, Clustering, Bias detection, NER, Information Extraction, Synthetic Data Generation.

Recognitions

- **Patent:** Filed with US PTO for Generative Search For Text Documents for Bing Generative Search.
- **Winner:** Microsoft WEB XT Hackathon, 2nd prize winner for best hack implementing Edge Split Screen Project.
- **Award:** Best individual contributor award from team manager at IP Infusion for BGP labelled Unicast.
- **Award:** Best individual contributor award from CEO of IP Infusion for flawless swift implementation of IS-IS Loop Free Alternate- IP Fast Reroute.
- **Award:** SNAP award from manager at IP Infusion for VRF implementation that led to winning Apple evaluation.
- **Bonus:** Retention bonus of > 10,00,000 INR from IP Infusion for top talent.