

Yang, Bo

Machine Learning Engineer

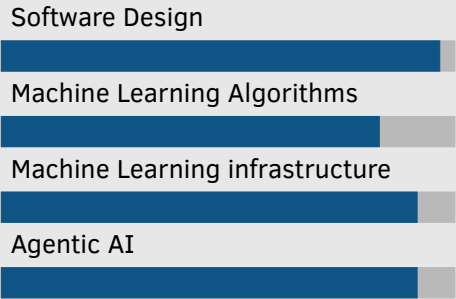
- +1-415-420-9693
- https://yang-bo.com
- yang-bo@yang-bo.com
- 955091
- Atry
- yang-bo

About me

I operate as half researcher, half engineer. The researcher in me creates and maintains open-source projects (3000+ stars) and publications in deep learning and functional programming. The engineer in me architects production systems from scratch for ambiguous domains, building the best practices for serverless LLM platforms and multi-agent systems where none existed before.

This duality defines my technical expertise: I am half systems programmer, half AI architect. I leverage a deep foundation in programming frameworks and performance engineering to build and operationalize the next frontier of AI.

Skills



Scala ★5.5 Python ★5.5 Hack ★5.5

Haxe ★5.5 OpenCL ★5 DevOps ★5

Agile Project Management ★4.3

ActionScript ★4 C++ ★4 Java ★4

JavaScript ★4 Rust ★4 OCaml ★4

Testing ★3 C ★3 Spark ★3

Akka ★3

(*)[The skill scale is from 0 (Fundamental Awareness) to 6 (Expert).]

Experience

| | | |
|-----------|--|----------------------------|
| 2024- | Tacnode | Machine Learning Engineer |
| | Built a serverless platform for LLM fine-tuning, test-time training, and inference, integrated with Discord and Hugging Face. Developed a multi-agent agentic system for database engineering. Owned LLMOps/MLOps, DevOps, and observability end-to-end. | |
| 2023-2024 | Gradient.AI | Machine Learning Engineer |
| | Designed and shipped serverless fine-tuning and inference API services for large language models (LLMs), enabling elastic scaling in a multi-tenant environment. | |
| 2019-2022 | Meta Platforms Inc | Software Developer |
| | Built machine-learning infrastructure for payments fraud detection. Maintained the HHVM open-source branch and Hack standard libraries, ported HHVM to multiple Linux distributions, and built its CI pipeline. | |
| 2015-2019 | Thoughtworks Inc | Lead / Senior Consultant |
| | Delivered product line engineering consulting for Huawei and led architecture for Scala projects for RCA. Founded and maintained open-source projects including Binding.scala and DeepLearning.scala. | |
| 2006-2014 | Various Game Companies | Game Developer / Tech Lead |
| | Developed and led teams to build multiple game engines, mobile/web/social games, and 3D applications. Key contributions include engine architecture, development tools, CI/CD pipelines, and open-source libraries. | |

Open-source projects

| | | |
|------------|--|----------------------------------|
| 2021-2022 | HHVM | C++/Rust/OCaml |
| | is Meta's virtual machine to run Hack. I ported the open source branch of HHVM to NixOS and created CI pipeline for building, testing, and packaging HHVM. | |
| since 2018 | Dsl.scala | Scala |
| | is a framework to extend Scala language features in libraries. I also ported this framework to Haskell as Control.Dsl. | |
| since 2016 | DeepLearning.scala and Compute.scala | Scala |
| | are libraries for creating complex neural networks. | |
| since 2015 | Binding.scala | Scala |
| | is a reactive templating language for creating web and desktop GUI. | |
| 2015-2016 | Microbuilder | Haxe / Scala / Java / JavaScript |
| | is a toolkit to build system across micro-services implemented in various languages communicating via RESTful JSON API. | |
| since 2012 | haxe-continuation | Haxe |
| | provides <code>async/await</code> syntax for Haxe. | |
| since 2010 | protoc-gen-as3 | ActionScript3 |
| | is a Protocol Buffers plugin for ActionScript 3. | |
| since 2007 | other 100+ of open-source projects | |

Publications

| | | |
|------|--|--------------|
| 2025 | CE-U: Cross Entropy Unlearning | first author |
| 2017 | Mondic Deep Learning | first author |
| 2017 | Ad-hoc polymorphic delimited continuations | first author |
| 2009 | Adobe Flex 3: Training from the Source | translator |

Talks

| | | |
|------|---|---------------------------|
| 2017 | Deeplearning.scala - Thoughts and practices on OSS AI framework | DBTC |
| 2017 | Monadic Deep Learning | Strange Loop |
| 2017 | Neural Networks & Functional Programming | ArchData Summit · Beijing |
| 2016 | More than Async | QCon · Beijing |
| 2015 | Specific-domain extension in an universal language | ECUG Con |

Education

| | | |
|-----------|---|------------------|
| 2002-2006 | Southwest University of Political Science and Law | Bachelor Of Laws |
|-----------|---|------------------|