

# ML + WebAssembly Working Group

3 October 2023

Andrew Brown

# Agenda

- Introductions
- Goals, structure
- High-level “problems”
- Discuss: other problems, requirements, use cases

# Introductions

- In one minute, please state:
  - your name
  - your company
  - your interest in wasi-nn

# Goals

- More community participation!
- Concretely, in the next six months:
  - move wasi-nn to phase 3
  - resolve key spec barriers to deployment
  - increase implementation robustness: more frameworks, more testing, more engines
  - move spec and implementations forward together
- Further out:
  - set the stage for future low-level extensions
  - ...

# Structure

- Proposed: an informal W3C work group that feeds into the WASI subgroup
- Some factors:
  - There's precedent for this with the wasi-crypto working group
  - Not all involved parties are BA members
  - BA SIG would be more implementation-focused; spec work needed
  - Keep administrative overhead low!
- Forum:
  - Video meeting every 3 weeks (notes—where?)
  - Review spec PRs on GitHub

# “Problems”

- Phase out WITX; progress with WIT (better language bindings!)
- Fix usability paper cuts ([names vs numbers](#), [resources](#), [errors](#), [execution target](#), etc.)
- Support LLMs ([bytes](#) vs string2tensor?)
- Unclear if immediately needed: metadata? alignment? model chaining?
- <your issue here>
- Question: which of these is a priority for you to contribute to?

# Not yet “problems”

- Training—wasi-nn is still focused on inference
- WebNN compatibility—or does someone want to work on this?
- Low-level API—prepare for this, but not needed in the next six months?
- Proposed: we keep these in mind but focus on the previous “problems” for the next six months (community building)

# Open Discussion

- To set up future meeting discussions:
  - Other use cases?
  - Other requirements?
  - Other critical problems?