

The graph displays the following metrics and their general trends:

- average-token-prob** (pink): Starts at ~0.4, increases to 1.0.
- verbalization-1s** (teal): Starts at ~0.0, increases to 1.0.
- verbalization-2s** (orange): Starts at ~0.2, increases to 1.0.
- p(true)** (red): Starts at ~0.3, increases to 1.0.
- trained-probe** (blue): Starts at ~0.1, increases to 1.0.
- perplexity** (purple): Starts at ~0.4, increases to 1.0.
- jaccard-degree** (tan): Starts at ~0.3, increases to 1.0.
- ood-probe** (grey): Starts at ~0.3, increases to 1.0.

