How to get more labeled training data?

Traditional Supervision: Have subject matter experts (SMEs) hand-label more training data

Too expensive!

Active Learning: Estimate which points are most valuable to solicit labels for

Semi-supervised Learning: Use structural assumptions to automatically leverage unlabeled data

Weak Supervision: Get lower-quality labels more efficiently and/or at a higher abstraction level

Get cheaper, lower-quality labels from non-experts

Get higher-level supervision over unlabeled data from SMEs

Use one or more (noisy/biased) pre-trained models to provide supervision

Transfer Learning: Use models already trained on a different task

Heuristics

Distant Supervision

Constraints

Expected distributions

Invariances