

Three advances that reinforced one another

Emmanuel Cassimatis · Part II · How Machines Became Intelligent

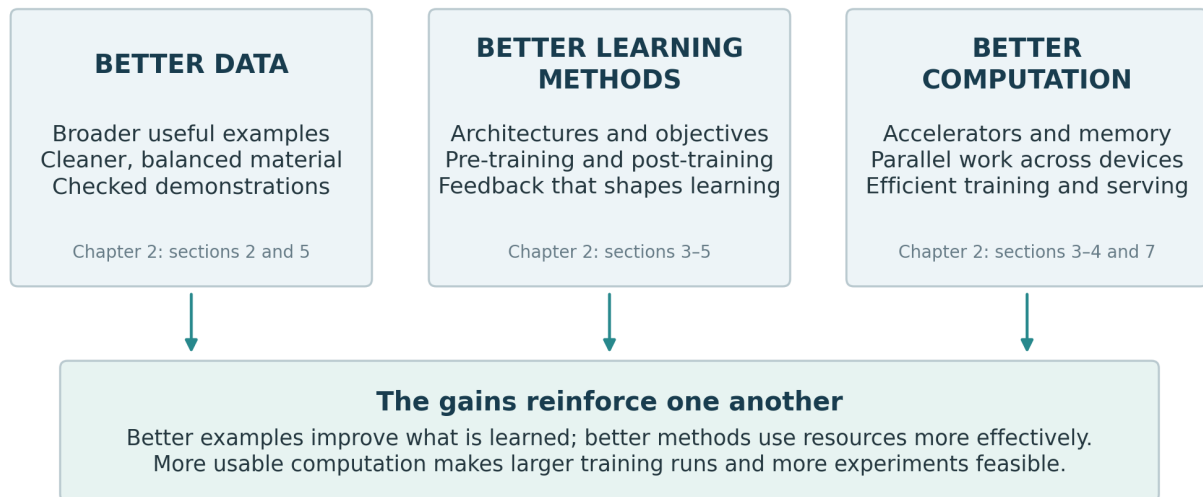


Figure II.2 Three interacting drivers of machine capability. Data determines the examples available; learning methods determine how experience changes the model; computation determines what can be carried out at a feasible scale. This is a conceptual framework, not a measured allocation of progress. The section references point to Chapter 2 of Part II. [II.2, II.5–II.7]

Better data

The supply of digital text, code and other material widened what machines could encounter. The gain also came from choosing that material more carefully: recovering a table correctly, removing repeated pages, balancing languages and checking demonstrations. FineWeb shows why preparation deserves attention alongside volume. For the reader building a model, the question is what each example teaches. In Chapter 2, section 2 follows that work; section 5 returns to data designed to teach an assistant how to respond. [II.5, II.6]

Better learning methods

New architectures and training objectives made more of those examples useful. The Transformer changed how a model relates positions in a sequence and made training easier to parallelize. Broad pre-training, instruction examples and feedback then supplied different kinds of learning signal. These are distinct advances in how a model is built and taught. Sections 3–5 of Chapter 2 explain them, including why mixture of experts is an architectural choice and preference learning is a training method. [II.2, II.8–II.10]

Better computation

Accelerators, memory, fast connections between devices and better software made larger experiments practical. Progress includes using the available hardware more effectively, as well as adding more of it. The balance matters: the Chinchilla experiments showed that changing how a fixed compute budget was divided between model size and training data could improve results. In Chapter 2, sections 3–4 explain the training constraints; section 7 follows the cost of serving answers. [II.6, II.7, II.24]

The boundaries overlap. An efficient learning method can reduce the computation required, while additional computation can make a different data mixture or a larger experiment possible. None of these drivers alone guarantees a reliable answer. Evaluation and oversight determine whether greater capability is useful in the intended setting. The historical sequence shows how these three advances eventually meet in the construction of one model.

Excerpt from the current book draft. Original figure and reference numbering retained. Selected passages are reproduced in order; layout has been adapted for this edition.

Research discussed: the Transformer, FineWeb, pre-training and post-training, and the Chinchilla compute-allocation experiments. The diagram is conceptual, not a measured attribution of progress.