- world war 2 (2017-2022) The transformer revolution
  - Transformers introduced by Google ("Attention is All You Need") – the atomic bomb of GenAI.
    - Before Transformers, models like RNNs (Recurrent Neural Networks) and LSTMs (Long Short-Term Memory) were used to process sequences (like sentences) 1 step at a time — sequentially.
    - That meant:
      - Word 1 → Word 2 → Word 3 → ... → Word N
      - Each word depends on the one before it.
      - This is slow and hard to parallelize, especially for long texts
    - Instead of recurrence, they use self-attention:
      - Each word in a sentence looks at every other word all at once.
      - The model figures out which words to pay attention to, no matter their position
    - The mother hit the daughter because she was drunk
  - Language Models Take Off
    - BERT (2018, Google):
      - Bidirectional understanding of language. Powers Google Search and beyond.
    - GPT-2 (2019, OpenAI):
      - Shockingly good at generating coherent text. OpenAI withholds full release initially due to misuse concerns.
    - T5 (Text-to-Text Transfer Transformer, Google, 2020)
      - Unified NLP framework: every NLP task becomes a “text-in, text-out” problem.
      - Transformers introduced by Google ("Attention is All You Need)
      - – the atomic bomb of GenAI.
  - 2020: OpenAI drops GPT-3 – shock and awe.
    - 175 billion parameters, trained on a massive corpus.
  - Text-to-image synthesis (DALL·E), Raw audio generation (Jukebox), Code generation (Codex)
  - Scale becomes the new weapon – billions of parameters, massive compute.