SMLab short talk on

LARGE CONCEPT MODELS

by
ADHILSHA ANSAD

05. 02. 2025

Stand on the shoulders of GIANTS

Large Concept Models:

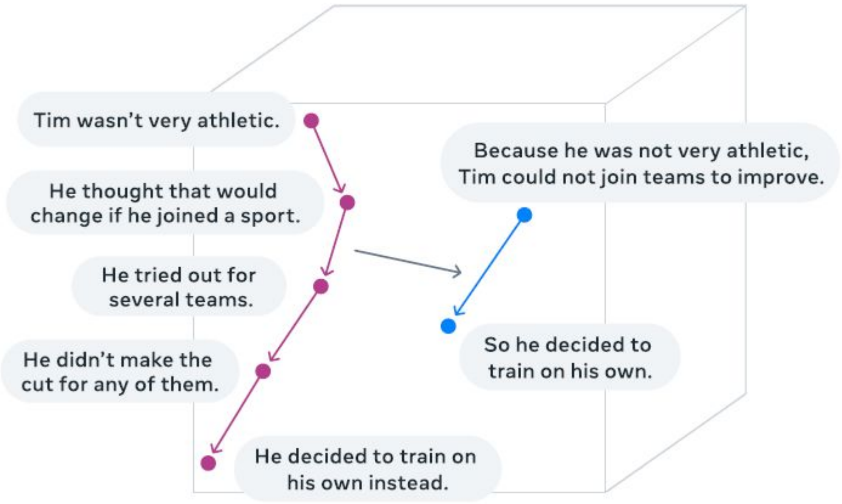
Language Modeling in a Sentence Representation Space

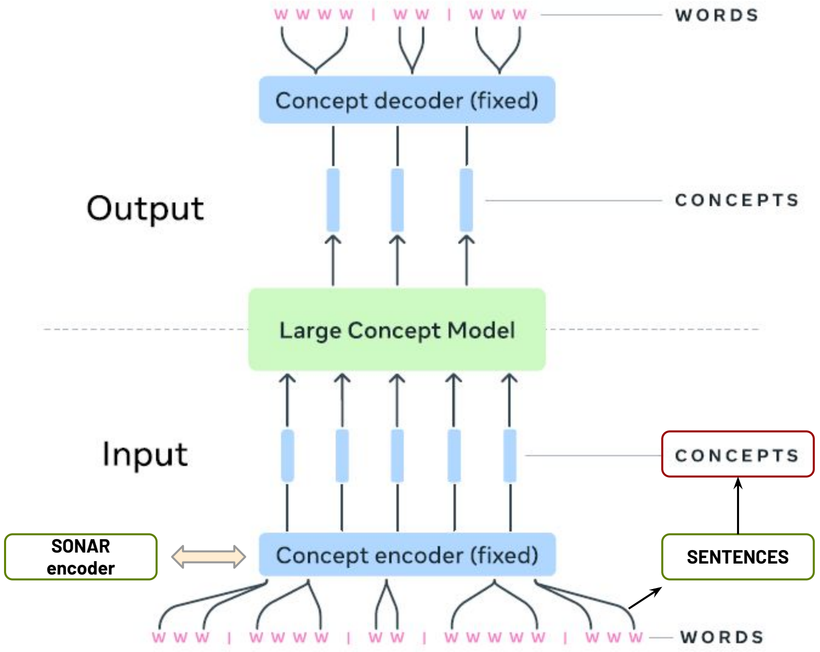
The LCM team, Loïc Barrault[*], Paul-Ambroise Duquenne[*], Maha Elbayad[*], Artyom Kozhevnikov[*], Belen Alastruey[†], Pierre Andrews[†], Mariano Coria[†], Guillaume Couairon[+†], Marta R. Costa-jussà[†], David Dale[†], Hady Elsahar[†], Kevin Heffernan[†], João Maria Janeiro[†], Tuan Tran[†], Christophe Ropers[†], Eduardo Sánchez[†], Robin San Roman[†], Alexandre Mourachko[‡], Safiyyah Saleem[‡], Holger Schwenk[‡]

FAIR at Meta

MOTIVATION

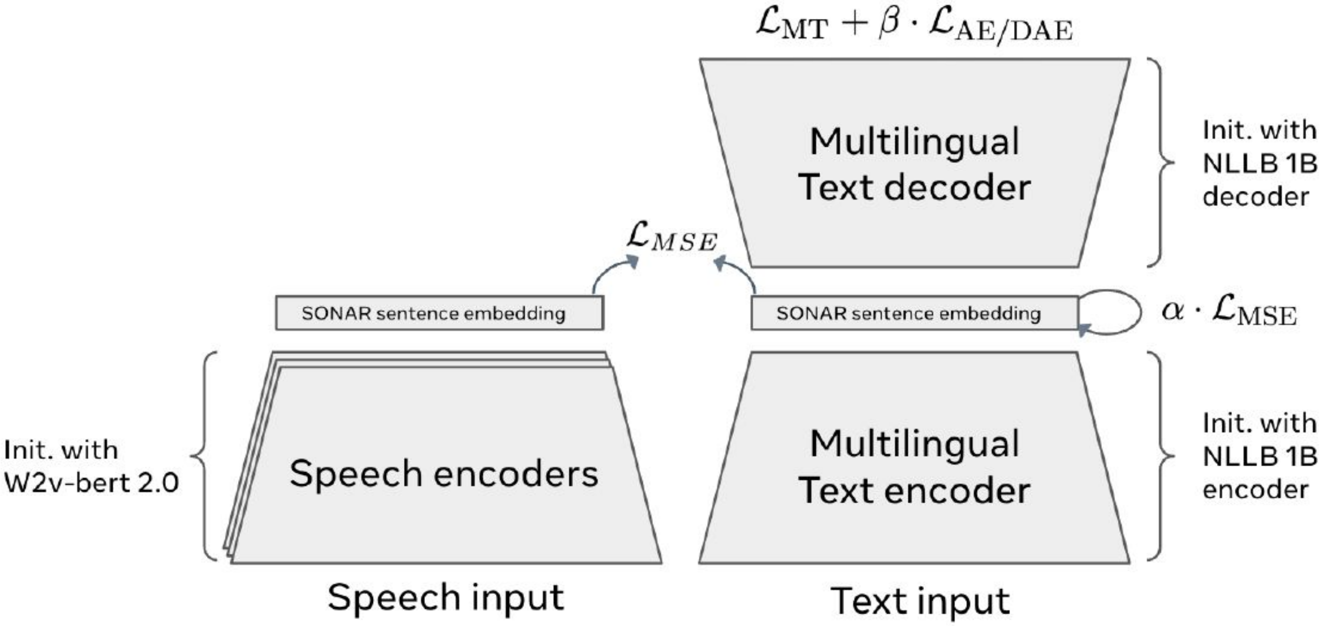
- Humans understand in different level of abstractions
- Input → hierarchical understanding → abstract ideas (concepts) → add details → Output
- Abstract Ideas - independent of language and modality





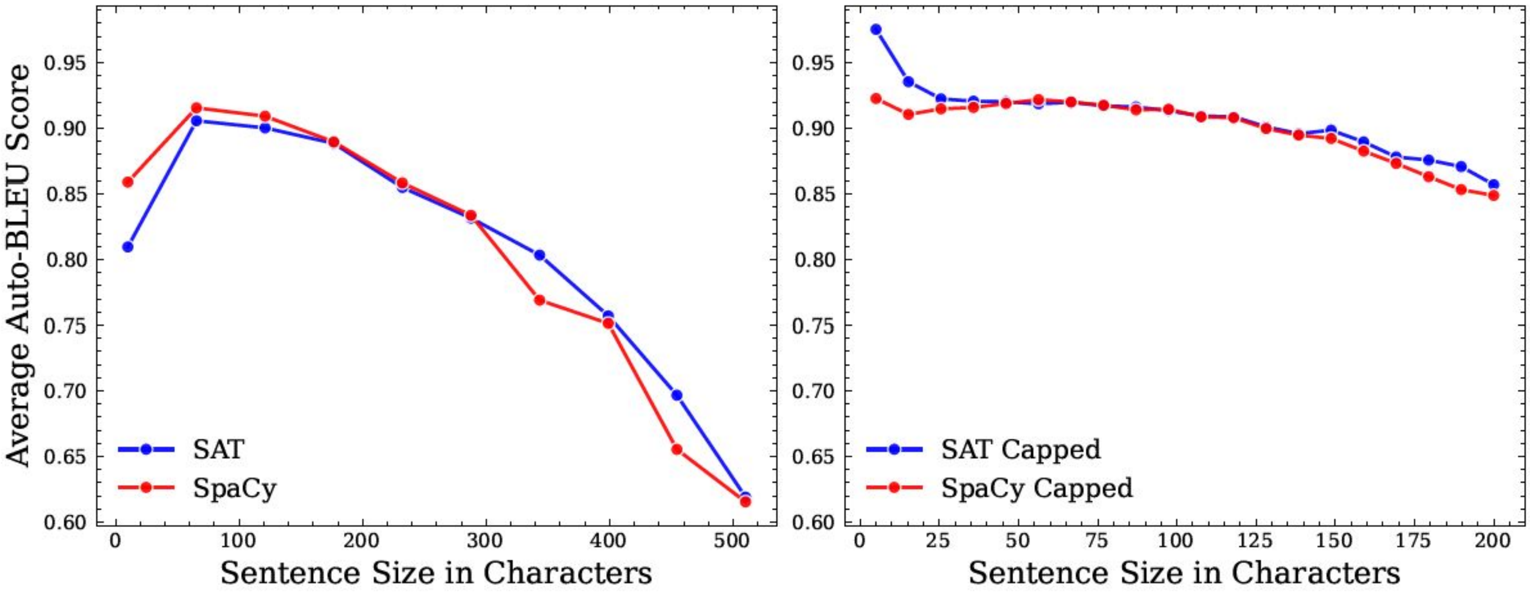
SONAR EMBEDDING SPACE

- encoder-decoder architecture
- Tasks/losses: Machine Translation (MT), Auto-encoding (AE) and denoising auto-encoding (DAE) (pre-training), Multi-lingual Representation learning (MSE loss), Representations for speech (MSE loss)
- SONAR sentence embedding used as *concepts* for LCM



TEXT SEGMENTATION

- SpaCy segmenter (rule-based approach) and SaT segmenter (predict sentence boundaries at the token level)
- maximum sentence length cap in characters - avoid long and complex sentences (SaT-Capped and SpaCy-Capped)
- AutoBLEU score between decoded text and the reference segment used as input.

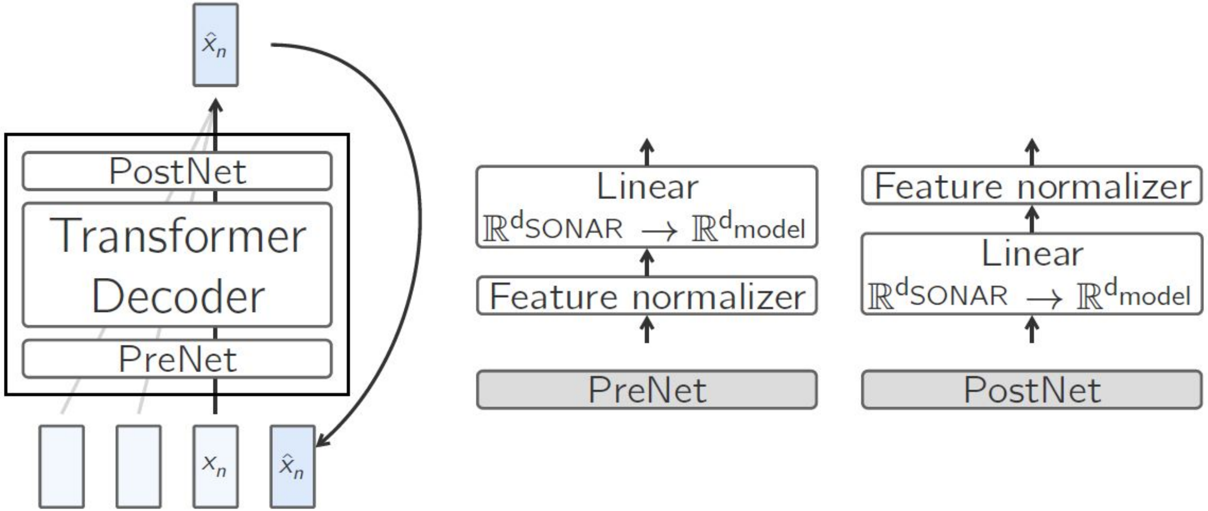


BASE-LCM

- decoder-only Transformer on next *concept* prediction

$$x^n = f(x < n; \theta) \quad \mathrm{MSE}(\hat{x}_n, x_n) = ||\hat{x}_n - x_n||^2$$

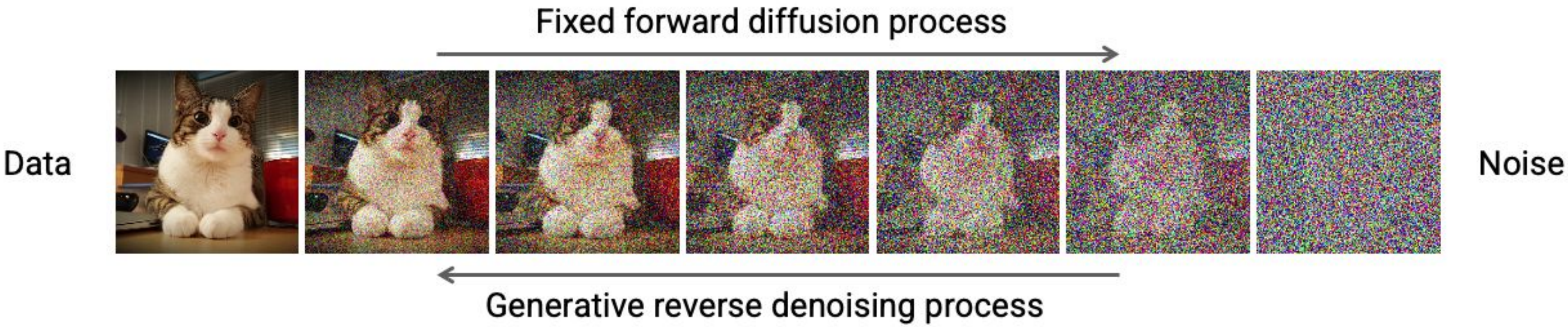
$$\mathcal{L}_{\text{BASE-LCM}}(\theta) = \mathbb{E}_{X \sim q}\left[\sum_{n=1}^{|x|} \mathrm{MSE}\left(f(x_{<n}; \theta), x_n\right)\right]$$



DIFFUSION MODELS

- Fixed forward diffusion by adding noise over several steps - noise schedules: Linear, Cosine, Quadratic etc.
- Generative reverse denoising process (NOTE: entire noise to be removed is predicted each step, but partially removed)
- New noise schedule: Sigmoid [from LCM]

$$\alpha_t^2 = \frac{f(t)}{f(0)}, \quad \text{where} \quad f(t) = \sigma(\delta - \gamma \cdot \text{logit}(t)) \qquad\qquad x_t = \alpha_t x_0 + \sigma_t \epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, I)$$

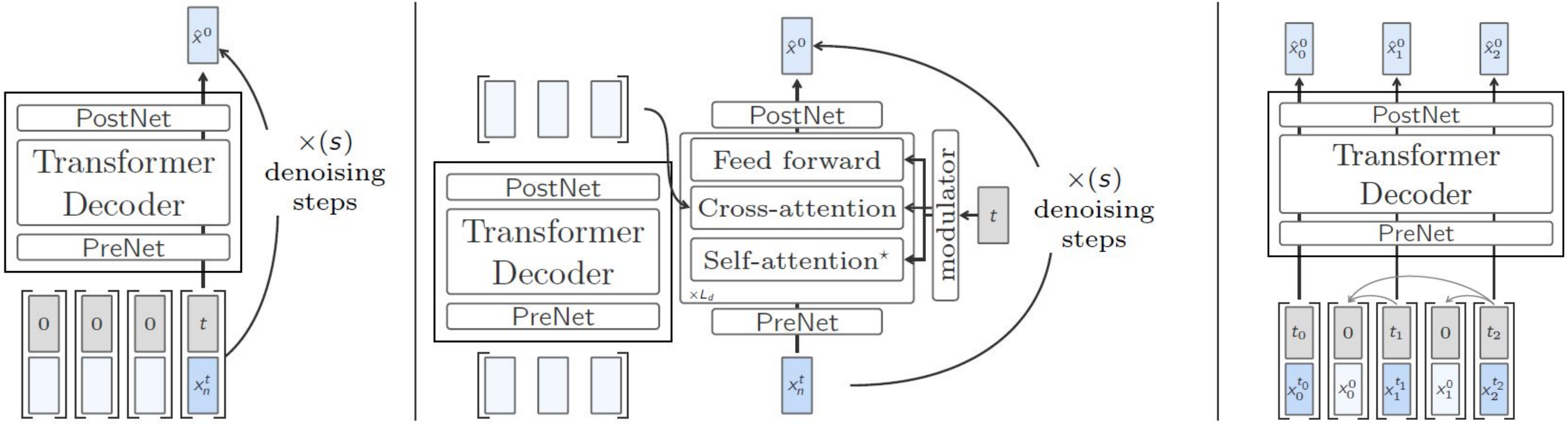


DIFFUSION-BASED LCMs

- Noisy input $x^t_{n'}$, previous clean sentence embeddings $x^0_{<n'} \rightarrow$ clean next sentence embedding x_0
- self-attention can be dropped with a certain probability for unconditional training.

$$\nabla_x \log_\gamma p(x|y) = (1-\gamma)\nabla_x \log p(x) + \gamma \nabla_x \log p(x|y)$$

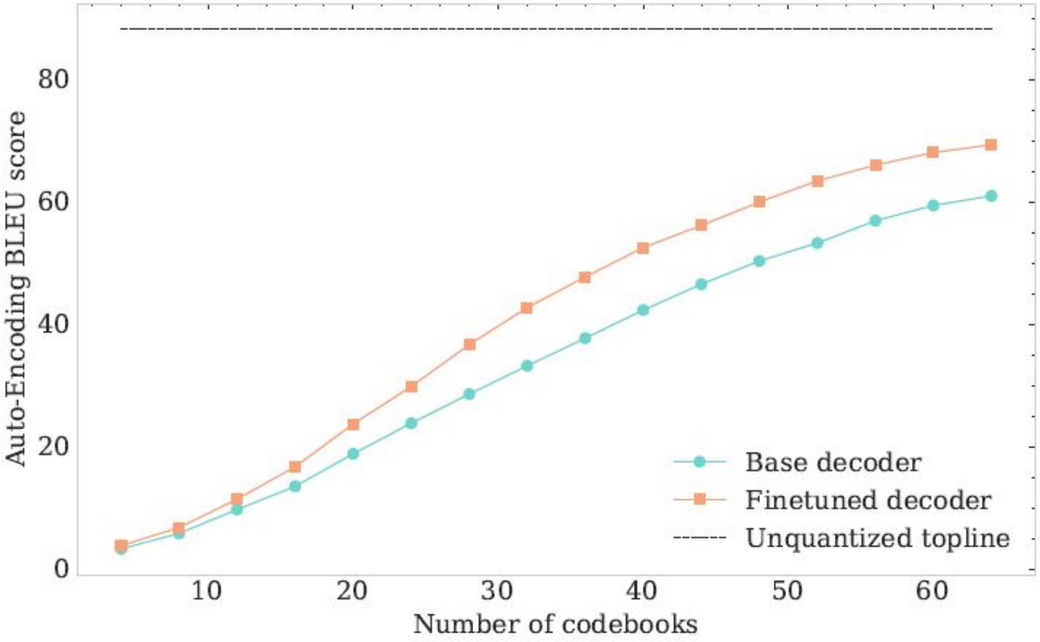
- Two-Tower LCM: contextualizer having casual self-attention, denoiser having cross attention (with encoded context) and Adaptive Layer Norm (AdaLN) modulator.



One-Tower LCM Two-Tower LCM Efficient Training

QUANTIZED SONAR EMBEDDINGS

- Quantization of SONAR space: Residual Vector Quantization (RVQ)
  - continuous input embeddings → nearest entry in a learnt codebook
  - iteratively quantize residual errors using additional codebooks
- Finetuning the SONAR decoder on quantized representations.
  - Also on residual representations from intermediate codebooks with probability p=0.3



QUANTIZED LCM

- Compared to diffusion LCM,
  - Denoising task X ; iterative generation of SONAR embeddings based on intermediate quantized representations √
  - input representations → intermediate quantized representations
  - diffusion timestep embeddings → codebook index embeddings
  - intermediate representation: 0 += predicted residual centroid embeddings , iteratively for all codebooks

- Quant-LCM-D training to predict the unit from the next codebook
  - $n_{\text{codebooks}} \cdot n_{\text{units-per-codebook}}$ outputs X ; $n_{\text{units-per-codebook}}$ outputs √
  - given cumulative sum of centroid embeddings of the first *k–1* codebooks as input
  - unit from codebook *k* of the target embedding as target index for cross entropy loss computation
- Quant-LCM-C training to predict residual embedding

CONTRIBUTIONS / ADVANTAGES

Language- and Modality-agnostic

- Cam encompass underlying reasoning process.
- Can be trained on all languages and modalities at once, hence unbiased.

Long context and long-form output

- Works with shorter 'tokens' compared to vanilla attention with quadratic complexity

Unparalleled zero-shot generalization

- Can be applied to any language and modality supported by the SONAR encoders, independent of training.

Modularity and extensibility

- New languages or modalities can be easily added.
- Concept encoders and decoders can be independently developed and optimized, bypassing modality competition.

LIMITATIONS AND FUTURE RESEARCH

Choice of the embedding space

- Possible research into other Language and modality agnostic embedding spaces, even beyond sentences.

Concept granularity

- Possible one-to-many mapping of sentence and concepts, to encompass complexity.
- Alternative to 'Prediction' of sentences since the sentence space is virtually unlimited.

Continuous versus discrete

- Build representation space for text capable of using modeling power of Diffusion model on continuous data.

Modularity

- Define Concepts in image space as well as how to predict and use it.

REFERENCES

Papers

- LCM team, Barrault, L., Duquenne, P.-A., Elbayad, M., Kozhevnikov, A., Alastruey, B., Andrews, P., Coria, M., Couairon, G., Costa-jussà, M. R., Dale, D., Elsahar, H., Heffernan, K., Janeiro, J. M., Tran, T., Ropers, C., Sánchez, E., San Roman, R., Mourachko, A., Saleem, S., & Schwenk, H. (2024). ***Large Concept Models: Language Modeling in a Sentence Representation Space***. arXiv. arxiv.org/abs/2412.08821
- Duquenne, Paul-Ambroise, Holger Schwenk, and Benoit Sagot. ***SONAR: Sentence-Level Multimodal and Language-Agnostic Representations.*** arXiv, 2023, arxiv.org/abs/2308.11466.

Figures

- [Slides 3,5,6,8,9] - LCM paper
- [Slide 4] - SONAR paper
- [Slide 7] - Vahdat, A., & Kreis, K. (2022, April 26). Improving Diffusion Models as an Alternative To GANs, Part 1 | NVIDIA Technical Blog. NVIDIA Technical Blog.

Codes

- https://github.com/facebookresearch/large_concept_model
- https://github.com/facebookresearch/SONAR

THANK YOU…