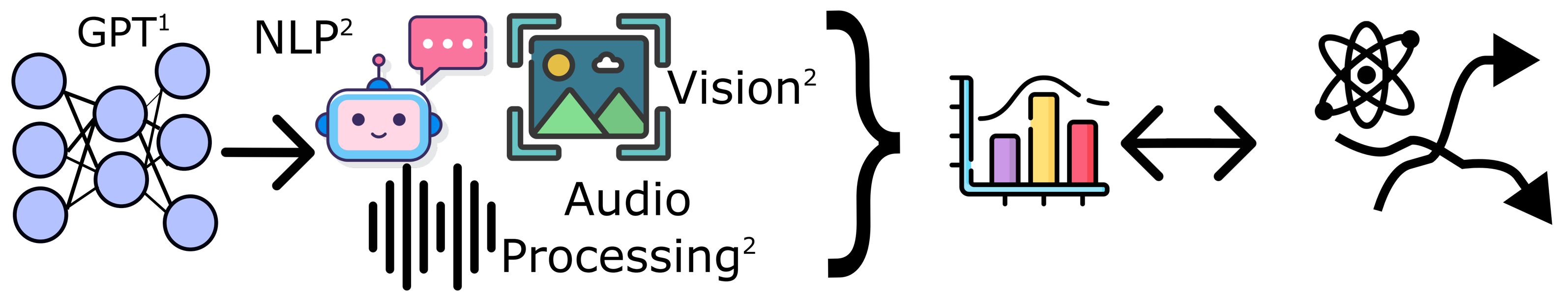
Transformers Networks for the Schrödinger Equation

Roberto Flórez Ablan[1], João Augusto Sobral, Mathias S. Scheurer
[1]rflorezablan@gmail.com

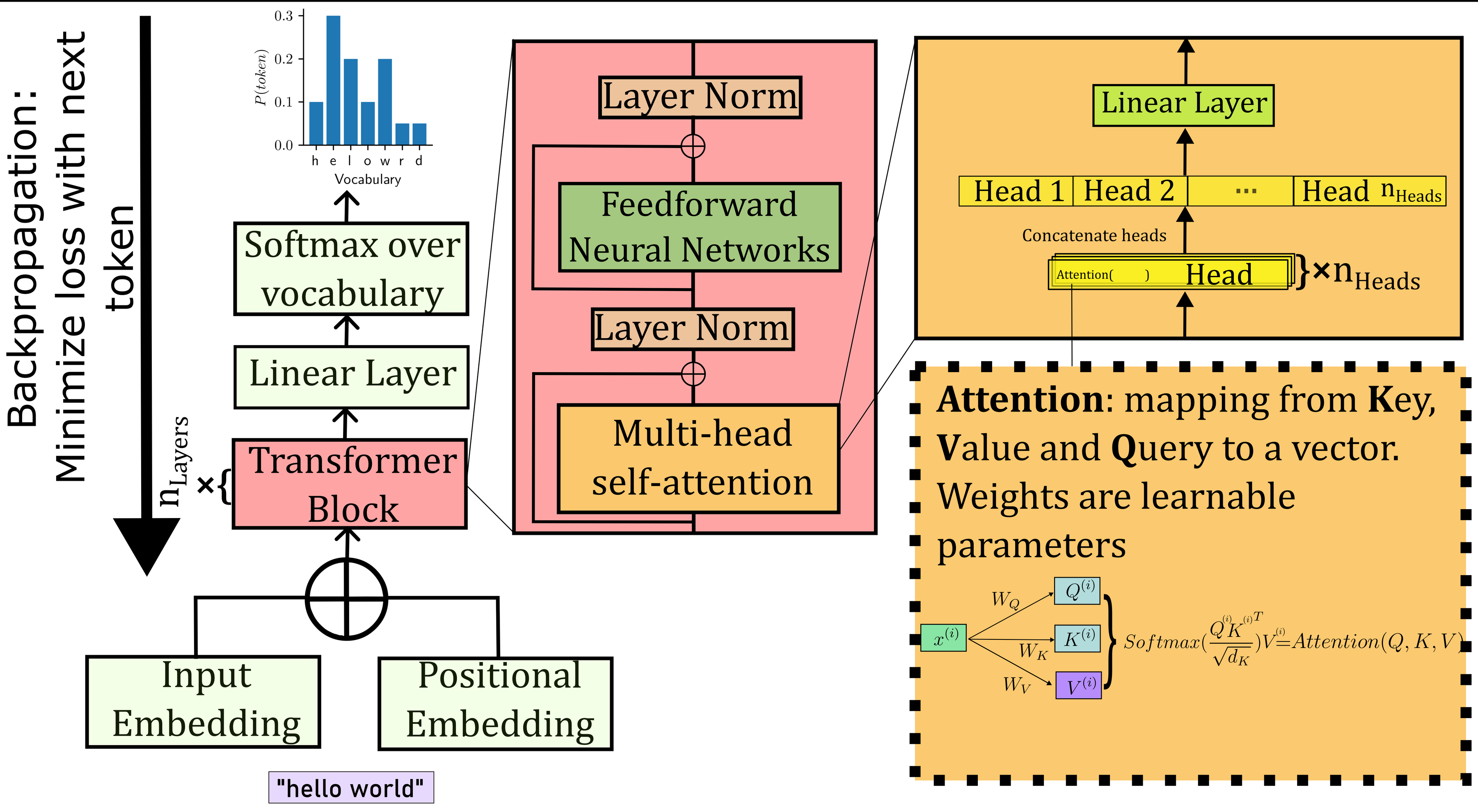
Motivation



- Transformers are a very sucessfull deep learning technique, especially in problems of stochastic nature. Quantum mechanical systems have this characteristic. Here we use them to study dynamics governed by the Schrödinger Equation.

[1]Vaswani et al., 2017; [2]Lin et al., 2021.

Transformer Archictecture



- A decoder-only transformer minimizes the loss function for next possible token.
- An input token is embedded in a n_{Emb} vector and the position information is included in a positional embedding, as attention is permutation equivariant.
- An implementation based on [3] was used in this work.

[3]Karpathy Andrej, 2022, NanoGPT, Github Repository.

Transformer Generation

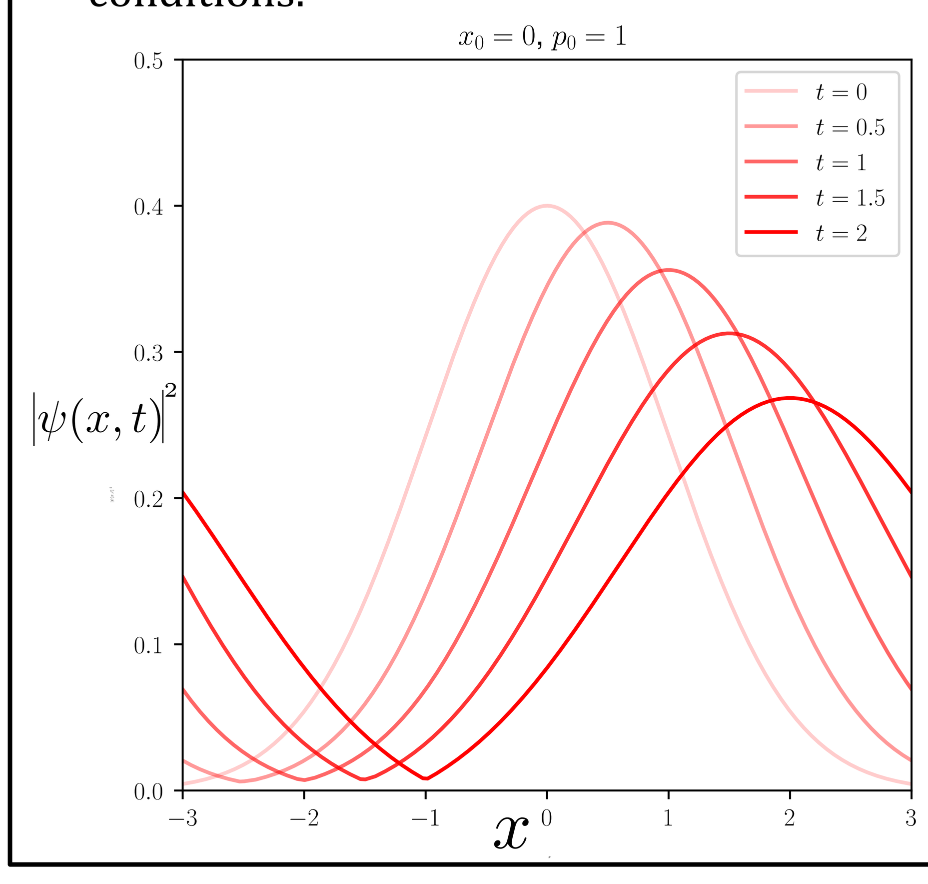
- As an initial benchmark of the implementation, we trained a model with $n_{Heads} = n_{Layers} = 4$ and $n_{Emb} = 64$, on a set of 20 Bossa Nova Songs.

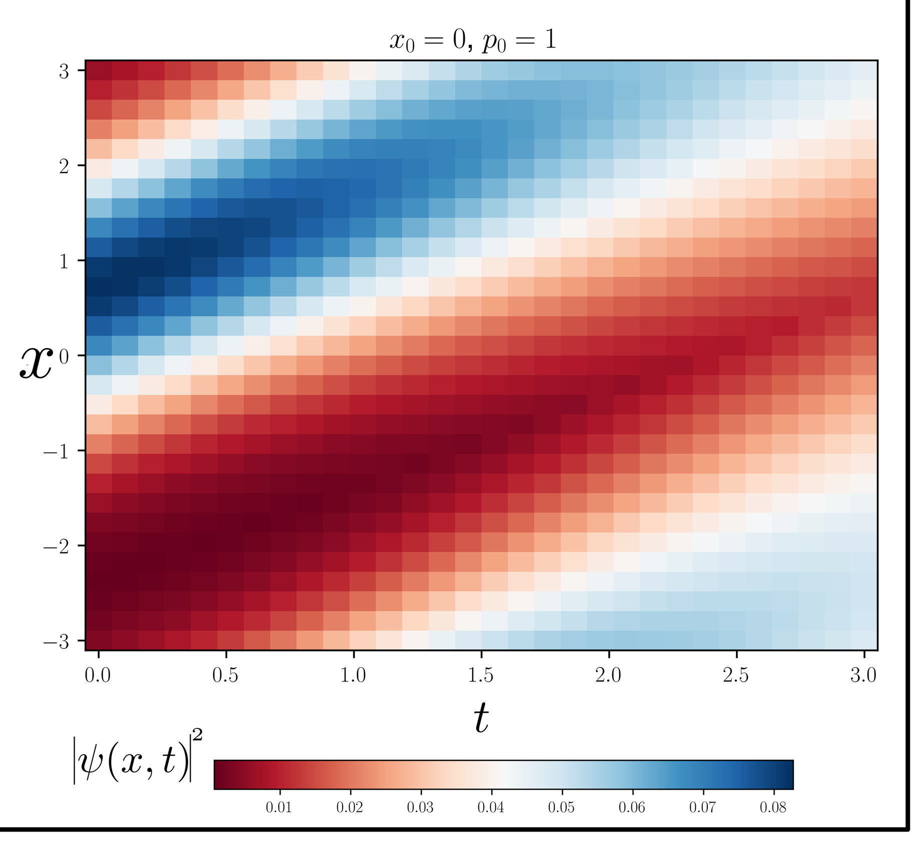
English human translation

- GPT understands Bossa Nova's structure in terms of paragraphs, sentences and authors.
- The lyrics are gibberish and with many portuguese gramatical mistakes.

Dynamics of a free quantum particle

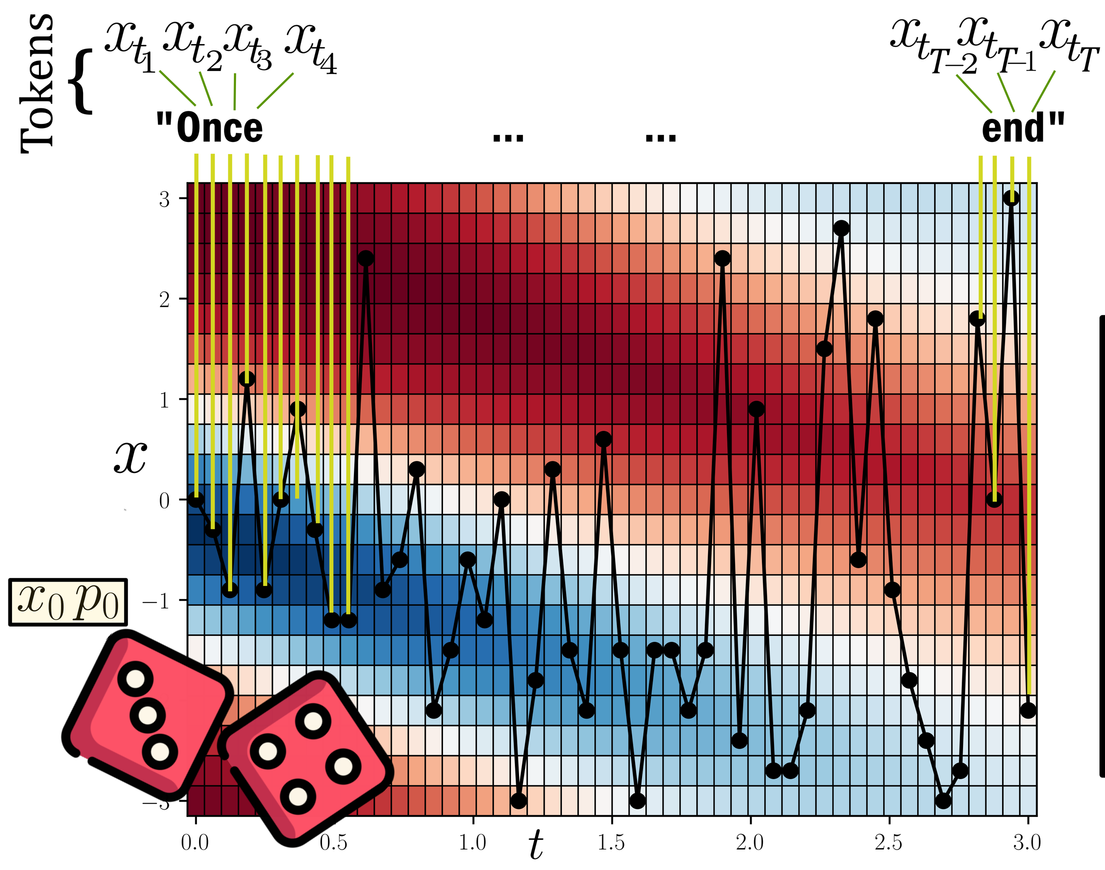
- In a quantum mechanical system:
  - The position of a particle is not uniquely determined. It is described by a wavefunction $\psi(x,t)$. The probability density of measuring a particle at position at time t is given by $|\psi(x,t)|^2$.
  - The evolution of $\psi(x,t)$ follows the Schrödinger equation: $i\hbar\frac{\partial\psi(x,t)}{\partial t}=\left[-\frac{\hbar^2}{2m}\frac{\partial}{\partial x^2}+V(x,t)\right]\psi(x,t)$
  - We implemented the time evolution of a free gaussian particle with periodic boundary conditions.





Trajectories dataset

- P different $|\psi(x,t)|^2$ were simulated with random pairs of x_0, p_0.
- K particle trajectories were sampled for each $|\psi(x,t)|^2$.
- Space and time were discretized in L (T) spacesteps (timesteps). Each spacestep corresponds to a token . Each token position in a sentence corresponds to a time step. The transformer was trained with 6×10^7 tokens, from *P=11000* distributions and *K=200* trajectories.



Dataset example

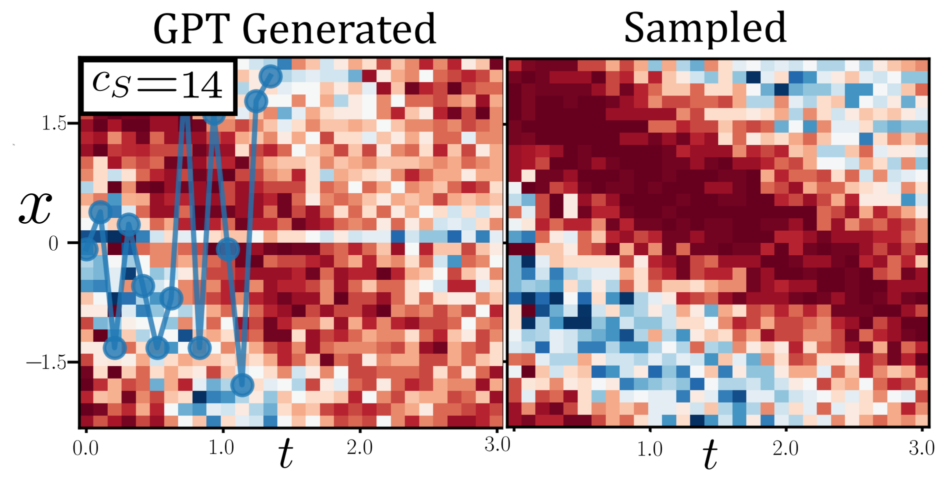
$$x_{1_{t_1}}\ x_{1_{t_2}}\ \dots\ x_{1_{t_T}}$$
$$x_{2_{t_1}}\ x_{2_{t_2}}\ \dots\ x_{2_{t_T}}$$
$$\vdots$$
$$x_{KP_{t_1}}\ x_{KP_{t_2}}\ \dots\ x_{KP_{t_T}}$$

Results: generating trajectories

- Ensembles of trajectories were generated with 3 "prompting" methods:
  - Method 1:
    - Randomly sampled positions $x_{rs_{t_i}}$ for the first c_S time steps. Transformer autocompletes trajectories and generates a new one. A histogram is built with the new trajectories.

Example: (c_S=3)

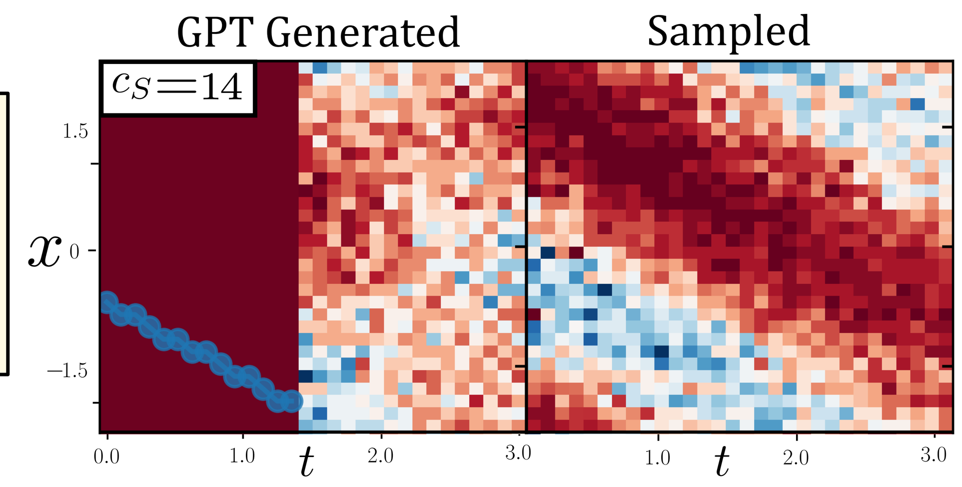
Prompt	Generation
$x_{rs_1}\ x_{rs_2}\ x_{rs_3}$	$x^{(0)}_{gpt_4}\ x^{(0)}_{gpt_5}\ \dots\ x^{(0)}_{gpt_T}$
$x^{(1)}_{gpt_1}\ x^{(1)}_{gpt_2}\ x^{(1)}_{gpt_3}$	$x^{(1)}_{gpt_4}\ x^{(1)}_{gpt_5}\ \dots\ x^{(1)}_{gpt_T}$
...	...



  - Method 2:
    - Most frequent positions x_{mf} for the first c_S time steps. Transformer autocompletes trajectories. A histogram is built from autocompleted trajectories.

Example: (c_S=3)

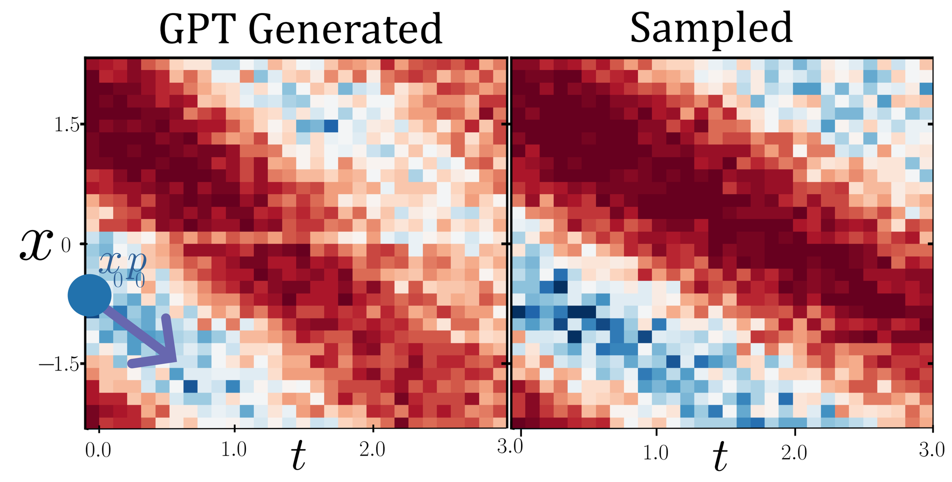
Prompt	Generation
$x_{mf_1}\ x_{mf_2}\ x_{mf_3}$	$x^{(0)}_{gpt_4}\ x^{(0)}_{gpt_5}\ \dots\ x^{(0)}_{gpt_T}$
$x_{mf_1}\ x_{mf_2}\ x_{mf_3}$	$x^{(1)}_{gpt_4}\ x^{(1)}_{gpt_5}\ \dots\ x^{(1)}_{gpt_T}$
...	...



  - Method 3:
    - Initial momentum and position were included in the dataset and given as prompt.

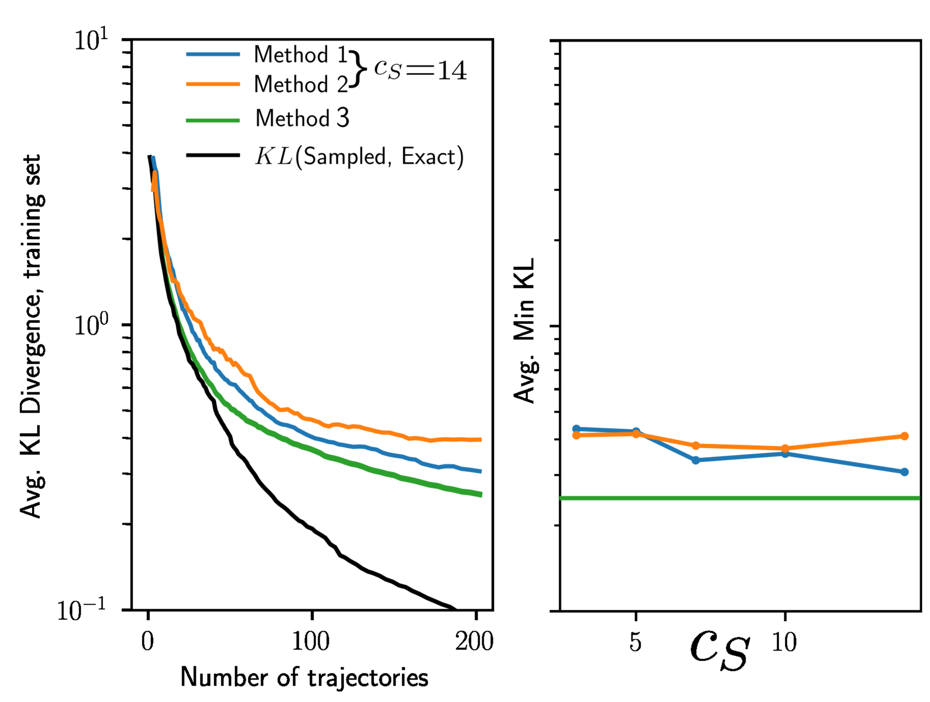
Example:

Prompt	Generation
$x_0 p_0;$	$x^{(0)}_{gpt_1}\ x^{(0)}_{gpt_2}\ \dots\ x^{(0)}_{gpt_T}$
$x_0 p_0;$	$x^{(1)}_{gpt_1}\ x^{(1)}_{gpt_2}\ \dots\ x^{(1)}_{gpt_T}$
...	...



- The resemblance between a transformer generated probability distribution and the exact probability distribution can be quantified with the Kullback–Leibler (KL) divergence.

$$KL(P||Q)=\sum_i P_i\log\frac{P_i}{Q_i}$$



Conclusions

- Transformers are able to partially reproduce distributions from quantum mechanics.
- Explicit information of initial momentum and position seems to be more powerful than context about trajectory positions. More simulations are necessary to understand the best prompting method.
- This proof of concept, is a first step to address harder problems in quantum mechanics, using the transformer architecture.