Large Language Models: from Drug Discovery to Clinical Trials
LANGUAGE MODELLING VIA NEURAL NETWORKS
What is an artificial neural network?
What is a deep neural network?

\[ a^{(1)} = \sigma \left( \sum_{i=1}^{n} w_{1,i} a^{(0)}_{i} + b^{(0)}_{1} \right) \]

\[ a^{(1)} = \sigma \left( \sum_{i=1}^{m} \left( \begin{array}{c} w_{1,0} a^{(0)}_{0} \\ w_{1,1} a^{(0)}_{1} \\ \vdots \\ w_{m,0} a^{(0)}_{m} \end{array} \right) \right) + \left( \begin{array}{c} b^{(0)}_{1} \\ b^{(0)}_{2} \\ \vdots \\ b^{(0)}_{m} \end{array} \right) \]

**Language model**

**Definition:** A *probabilistic model* that assigns a probability \( P(w_1, w_2, \ldots, w_n) \) to every finite word sequence \( w_1, \ldots, w_n \)

**Auto-regressive model**

\[
P(w_1, \ldots, w_n) = P(w_1) \times P(w_2 | w_1) \times P(w_3 | w_1, w_2) \times \ldots \times P(w_n | w_1, \ldots, w_{n-1})
\]

\[
P(w_1, \ldots, w_n) \approx P(w_1) \times P(w_2 | w_1) \times P(w_3 | w_2) \times \ldots \times P(w_n | w_{n-1})
\]

\[V = \{\text{chef, cooked, meal, the}\}\]

\[
P(\text{the, chef, cooked, the, meal}) = \]
\[
P(\text{the}) \times P(\text{chef | the}) \times P(\text{cooked | the chef}) \times \ldots \times P(\text{meal | the chef cooked the})
\]

\[
P(\text{the, chef, cooked, the, meal}) = 0.0200
\]
\[
P(\text{the, meal, cooked, the, chef}) = 0.0100
\]
\[
P(\text{chef, the, the, meal, cooked}) = 0.0001
\]

You shall know a word by the company it keeps!


GPT-3 still acts in this way, but the model is implemented as a very large neural network of 175-billion parameters!
Representation and generation

Source: ai.googleblog.com
# A brief story of (modern) LLMs

<table>
<thead>
<tr>
<th>Concept</th>
<th>Authors</th>
<th>Year</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLM</td>
<td>Bengio <em>et al.</em></td>
<td>2003</td>
<td>10k+</td>
</tr>
<tr>
<td>word2vec</td>
<td>Mikolov <em>et al.</em></td>
<td>2013</td>
<td>37k+</td>
</tr>
<tr>
<td>seq-to-seq</td>
<td>Sutskever <em>et al.</em></td>
<td>2014</td>
<td>23k+</td>
</tr>
<tr>
<td>Transformers</td>
<td>Vaswani <em>et al.</em></td>
<td>2017</td>
<td>86k+</td>
</tr>
<tr>
<td>ELMo</td>
<td>Peters <em>et al.</em></td>
<td>2018</td>
<td>13k+</td>
</tr>
<tr>
<td>BERT</td>
<td>Devlin <em>et al.</em></td>
<td>2018</td>
<td>76k+</td>
</tr>
<tr>
<td>GPT-2</td>
<td>Radford <em>et al.</em></td>
<td>2019</td>
<td>6k+</td>
</tr>
</tbody>
</table>

...  
GPT-3  
Claude  
ChatGPT  
GPT-4  
...
A brief story of LLMs

A Neural Probabilistic Language Model, Bengio et al., 2003

Efficient Estimation of Word Representations in Vector Space, Mikolov et al., 2013

Sequence to Sequence Learning with Neural Networks, Sutskever et al., 2014
A brief story of LLMs

Attention is all you need, Vaswani et al., 2017
A brief story of LLMs

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Devlin et al., 2018

Language Models are Unsupervised Multitask Learners, Radford et al., 2019
Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond
Yang et al., 2023

Scaling Laws for Neural Language Models,
Kaplan et al., 2020
FROM NATURAL TO CHEMICAL LANGUAGE
**Natural language**

<table>
<thead>
<tr>
<th>Character</th>
<th>a-zA-Z0-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>water</td>
</tr>
<tr>
<td>Sentence</td>
<td>I want water</td>
</tr>
</tbody>
</table>

**Chemical language**

<table>
<thead>
<tr>
<th>Atom</th>
<th>C, H, N, ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecule</td>
<td>H₂O</td>
</tr>
<tr>
<td>Reaction</td>
<td>2H₂ + O₂ → 2H₂O</td>
</tr>
</tbody>
</table>

Nc1sccc1C#N.O[N+] (=O)C1=C(F)C=C(F)C(F)=C1 >
clccoc1.[H-].[Na+] >
O[N+] (=O)C1=CC(F)=C(F)C=C1NC1SCCC1C#N

\[
\begin{align*}
\text{Nc1sccc1C#N.O[N+] (=O)C1=C(F)C=C(F)C(F)=C1 } & > \\
\text{clccoc1.[H-].[Na+] } & > \\
\text{O[N+] (=O)C1=CC(F)=C(F)C=C1NC1SCCC1C#N}
\end{align*}
\]
Molecular string representations

**SMILES**: simplified molecular-input line-entry system

**SELFIES**: SELF-referencing embedded string

*Branch/ring constraints*

**InChI**: International Chemical Identifier (Uniqueness)

SELFIES and the future of molecular string representations, Krenn et al., 2022.
Çizgiyi geçtin ! <eos> Over the line ! <eos>
Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction, Schwaler et al., 2019
Augmenting large language models with chemistry tools
Bran et al., 2023
APPLICATIONS OF LLMS IN DRUG DISCOVERY AND DEVELOPMENT
Transformer Performance for Chemical Reactions: Analysis of Different Predictive and Evaluation Scenarios
Jaume-Santero, Bornet et al., JCIM, 2023

SYNTHESIS AND ONE-STEP RETROSYNTHESIS PREDICTION
\[ \text{Reactants} \quad \text{Reagents} \quad \xrightarrow{\text{Molecular Transformer}} \quad \text{Product} \]

\[ \text{Nc1scccc1C#N.O[N+] (}=O)\text{C1}=\text{C(F)}\text{C}=\text{C(F)}\text{C}(\text{F})=\text{C1} > \text{c1ccoc1.[H-].[Na+] } \]

<table>
<thead>
<tr>
<th>Score</th>
<th>SMILES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>N#C1ccsc1Nc1cc(F)c(F)ccl[N+] (=O) [O-]</td>
</tr>
<tr>
<td>7.3x10^{-6}</td>
<td>N#C1ccsc1Nc1cc(F)c(F)ccl[N+] (=O) [O-]</td>
</tr>
<tr>
<td>4.6x10^{-6}</td>
<td>N#C1ccsc1Nc1cc(F)c([N+] (=O) [O-]) cc1F</td>
</tr>
<tr>
<td>4.2x10^{-6}</td>
<td>N#C1ccsc1Nc1cc(F)c(N)ccl[N+] (=O) [O-]</td>
</tr>
</tbody>
</table>
The diagram illustrates the process of predicting products and reactants in reactions. The model, denoted as $P_{\text{model}}$, predicts a product $P$ and a reactant $\hat{R}_c$. The model's predictions are then compared to the gold standard to assess accuracy.

The graph on the right shows the roundtrip accuracy of predictions with and without reagents, plotted against data augmentation levels from $x1$ to $x20$. The accuracy is measured on the y-axis, which ranges from 0.2 to 1.0.

The bar chart compares the number of reactions annotated correctly by Ph.D. students and M.Sc. students. The reactions are categorized into three types: wrong, semi-correct, and correct. The chart shows a significant difference between the two groups, with the Ph.D. students annotating more reactions correctly, as indicated by the higher bars in the correct category.
Named entity recognition in chemical patents using ensemble of contextual language models
Copara et al., CLEF, 2020

REACTION EXTRACTION FROM PATENTS
Example 15A

**REACTION_PRODUCT**

1-(2-Methoxyethyl)-5-methyl-2,4-dioxo-3-(2-phenylethyl)-1,2,3,4-tetrahydrothieno[2,3-d]pyrimidine-6-carboxylic acid

**STARTING_MATERIAL**

75 ml of trifluoroacetic acid were added to a solution of 5.0 g (11.2 mmol) of the compound from Ex. 10A in 225 ml of dichloromethane, and the mixture was stirred at RT for 2 h. The reaction mixture was then concentrated to dryness on a rotary evaporator. The remaining residue was stirred in diethyl ether and filtered off with suction, and the solid was dried under high vacuum. 4.1 g (92% of theory) of the title compound were obtained.

### Evaluation Table

<table>
<thead>
<tr>
<th>Team</th>
<th>Precision</th>
<th>Recall</th>
<th>F₁-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exact</td>
<td>relaxed</td>
<td>exact</td>
</tr>
<tr>
<td>Melaxtech</td>
<td>0.9571</td>
<td>0.9690</td>
<td>0.9570</td>
</tr>
<tr>
<td><strong>DS4DH (run 3)</strong></td>
<td>0.9378</td>
<td>0.9692</td>
<td>0.9087</td>
</tr>
<tr>
<td><strong>DS4DH (run 2)</strong></td>
<td>0.9083</td>
<td>0.9510</td>
<td>0.9114</td>
</tr>
<tr>
<td>Baseline (BANNER)</td>
<td>0.9071</td>
<td>0.9219</td>
<td>0.8723</td>
</tr>
</tbody>
</table>
Prédiction des effets indésirables potentiels des médicaments pour les nouvelles molécules en cours d'essais cliniques

Vicente Alvarez & Yazdani, 2022

PREDICTING POTENTIAL ADRS FOR NEW MOLECULES
Blood level
WHO-ATC
ADMET
…

Data sources

Integrated database

Predictive algorithm

Predicted SOC ADEs

Clinical trials

Drug Bank

MedDRA

Dataset

LLM

SOC 1
SOC 2
SOC 3
SOC 23
SOC 24
SOC 25
SOC 26
Dosage, route, criteria, … (Strings)

<table>
<thead>
<tr>
<th>Model</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>38%</td>
</tr>
<tr>
<td>SMILES (ChemBERTa)</td>
<td>58%</td>
</tr>
<tr>
<td>Baseline (CT+DB)</td>
<td>64%</td>
</tr>
<tr>
<td>Baseline + MACCS</td>
<td>64%</td>
</tr>
<tr>
<td>Baseline + GROVER</td>
<td>63%</td>
</tr>
<tr>
<td>Baseline + ESMFold-2</td>
<td>64%</td>
</tr>
<tr>
<td>Dosage/route + SMILES (ChemBERTa)</td>
<td>69%</td>
</tr>
</tbody>
</table>
Deep learning-based risk prediction for interventional clinical trials based on protocol design: A retrospective study
Ferdowsi et al., Patterns, 2023

RISK PREDICTION OF CLINICAL TRIALS
COVID-19

Protocol

NCT04372602

Eligibility

Inclusion

1. Age ≥ 18 years or older
2. A diagnosis of advanced COVID-19
3. Adequate hematologic function

Exclusion

1. Known allergy
2. Pregnant and/or breastfeeding

Design

Phase

II

Primary

Overall Survival Through 28 Days

Secondary

Length of ICU stay...

Enrolment

28

Outcome

Status

Completed

Intervention

0

Duvelisib

Placebo

Peripheral blood draw
<table>
<thead>
<tr>
<th>Method</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1-score (%)</th>
<th>Accuracy (%)</th>
<th>AUROC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Elkin and Zhu, 2021)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>72.8</td>
</tr>
<tr>
<td>fastText (Joulin et al., 2017)</td>
<td>84.9</td>
<td>72.0</td>
<td>75.3</td>
<td>84.0</td>
<td>84.6</td>
</tr>
<tr>
<td>BERT-flat-1</td>
<td>65.1</td>
<td>67.6</td>
<td>62.6</td>
<td>63.5</td>
<td>72.5</td>
</tr>
<tr>
<td>BOW-flat-9</td>
<td>64.9</td>
<td>67.1</td>
<td>64.9</td>
<td>63.5</td>
<td>73.7</td>
</tr>
<tr>
<td>BERT-flat-9</td>
<td>81.4</td>
<td>81.4</td>
<td>81.4</td>
<td>84.2</td>
<td>89.1</td>
</tr>
<tr>
<td>BOW+RP-GCN-global</td>
<td>81.8</td>
<td>82.3</td>
<td>82.1</td>
<td>84.6</td>
<td>91.2</td>
</tr>
<tr>
<td>BERT-GCN-global</td>
<td>84.3</td>
<td>85.0</td>
<td>84.6</td>
<td>86.7</td>
<td>88.8</td>
</tr>
<tr>
<td>BOW+RP-GCN-selective-9</td>
<td>84.2</td>
<td>83.4</td>
<td>83.8</td>
<td>86.3</td>
<td>90.8</td>
</tr>
<tr>
<td>BERT-GCN-selective-9</td>
<td><strong>84.5</strong></td>
<td><strong>85.2</strong></td>
<td><strong>84.8</strong></td>
<td><strong>87.0</strong></td>
<td><strong>92.7</strong></td>
</tr>
</tbody>
</table>
IDENTIFYING ADRS IN SOCIAL NETWORK

DS4DH at #SMM4H 2023: Zero-Shot Adverse Drug Events Normalization using Sentence Transformers and Reciprocal-Rank Fusion
Yazdani et al., SMM4H, 2023
Paxlovid made my mouth taste bad, but it was worth it as my test was negative after 3 days.
Paxlovid made my mouth taste bad, but it was worth it as my test was negative after 3 days.
Named entity recognition

Ontology mapping

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall Precision</th>
<th>Overall Recall</th>
<th>Overall F1-score</th>
<th>Unseen Precision</th>
<th>Unseen Recall</th>
<th>Unseen F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.293</td>
<td>0.422</td>
<td>0.329</td>
<td>0.151</td>
<td>0.360</td>
<td>0.202</td>
</tr>
<tr>
<td>Median</td>
<td>0.249</td>
<td>0.405</td>
<td>0.322</td>
<td>0.128</td>
<td>0.354</td>
<td>0.195</td>
</tr>
<tr>
<td>Ours</td>
<td>0.449</td>
<td>0.405</td>
<td>0.426</td>
<td>0.249</td>
<td>0.354</td>
<td>0.292</td>
</tr>
</tbody>
</table>
Summary

- LLMs are powerful, generalisable machine learning models

- Drug discovery and development can benefit of the effectiveness LLMs in many ways
  - Synthesis/retrosynthesis, ADR prediction/extraction, CT analyses, …

- Outlook: While discriminative LLMs can provide strong performance in specific tasks, generative models can tackle more complex, general tasks
References


