

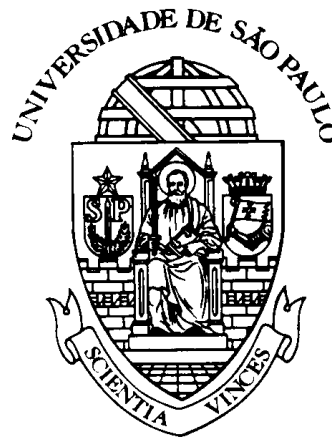
Aplicações de Inteligência Artificial em Pesquisa Científica

Guilherme Menegon Arantes

*Instituto de Estudos Avançados & Instituto de Química
Universidade de São Paulo*

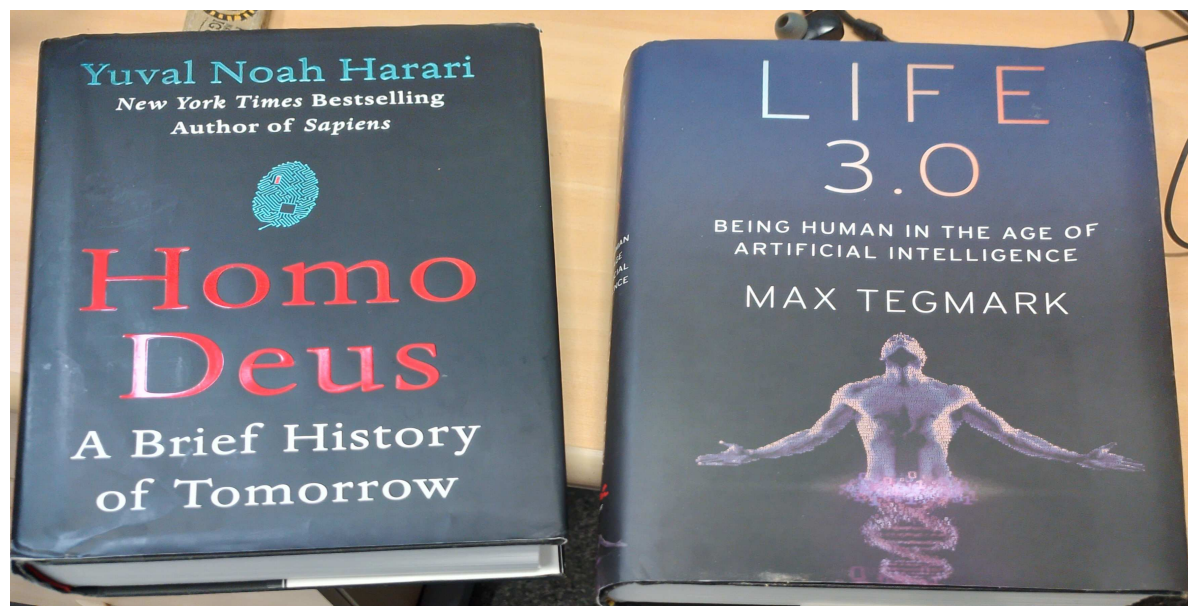
garantes@iq.usp.br

<http://gaznevada.iq.usp.br>



Não é nosso objetivo discutir:

- Superinteligência ou singularidade, cenários transcendentais ou apocalípticos
- Ciência interdisciplinar para compreender/modelar cognição



Objetivo aqui é pragmático

- Discutir IA como uma tecnologia ou ferramenta. Será (já é) uma revolução para cognição e/ou expressão humana, como a Internet foi para nossa comunicação
- Sugestão de leitura:
IA Ethics, Mark Coeckelbergh
<https://artificialintelligenceact.eu>



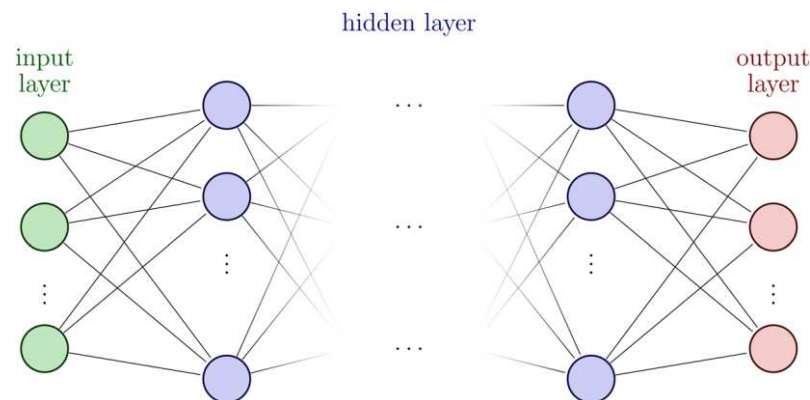
O que é IA?

- IA = Math (álgebra linear + estatística) + Dados + HPC

$y = f(x)$ Uma função é um mapeamento...

$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$ Mapeamento linear

$$\mathbf{y} = \mathbf{W}_{i+1} \sigma(\mathbf{W}_i \mathbf{x} + \mathbf{b}_i) + \mathbf{b}_{i+1}$$



Tipos de IA:

- IA simbólica (árvores de decisão, flowchart)
- IA não-profunda
 - Redução de dimensão, regressão
 - PCA, clusterização, SVM, etc.
- Redes neurais (NN) / Aprendizado de máquina (ML)
 - Deep Learning (DL): muitos NN em camadas
 - Várias arquiteturas: Convolution NN, Transformers, ...
 - Modelos generativos (DALL-E, ...)
 - Large Language Models (LLM)



LLM - Usos atuais:

- Gerar esboço de texto inicial
- Corrigir escrita (ortografia/sintaxe + gramática + sentido)
- Ajuda em programação simples
- Análise e publicação de dados científicos → mal-uso (dada complexidade), bias implícitos... Necessidade de melhores práticas na reportagem



LLM - Temores & Cuidados:

- Não é o novo livro. Auto-reciclagem de textos...
- Brecha de confidencialidade
- Recomendação àqueles que se sintam ameaçados pelo uso de ChatGPT *et al.*: Use-os em uma tarefa como “gere um texto com IA e em seguida avalie o texto gerado criticamente”. Isso ajudará os alunos a entender como as IAs funcionam e a desenvolver seu pensamento crítico.



Aplicações de IA para pesquisa em biomedicina

- Diagnóstico médico: Uso de IA em imagens médicas (radiografias, ressonâncias) e patologias.
- Análise de dados biomédicos: Genômica, proteômica e big data.
- Descoberta de fármacos: IA acelera a identificação de novos medicamentos.
- Modelagem e simulações computacionais de mecanismos biomoleculares



Mas antes, algumas informações sobre biomoléculas...



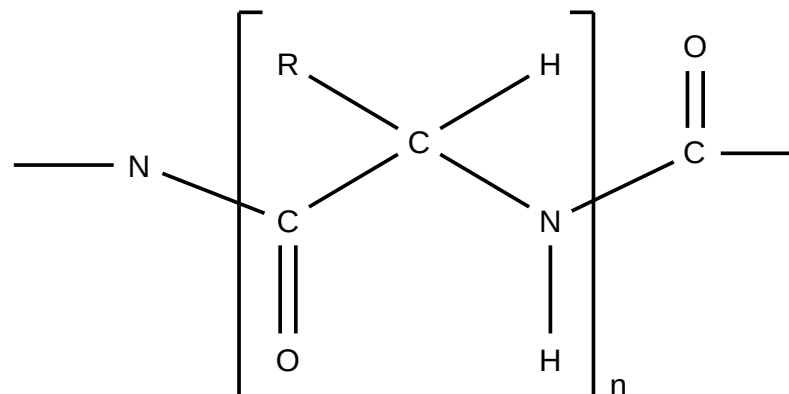
Dogma central do comportamento biomolecular

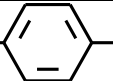
Estrutura → Mecanismo → Função

- Composição e forças intermoleculares → estrutura
- Temperatura → flutuações
- Estrutura + flutuações (selecionadas pela evolução) → função



Proteínas são compostos de aminoácidos

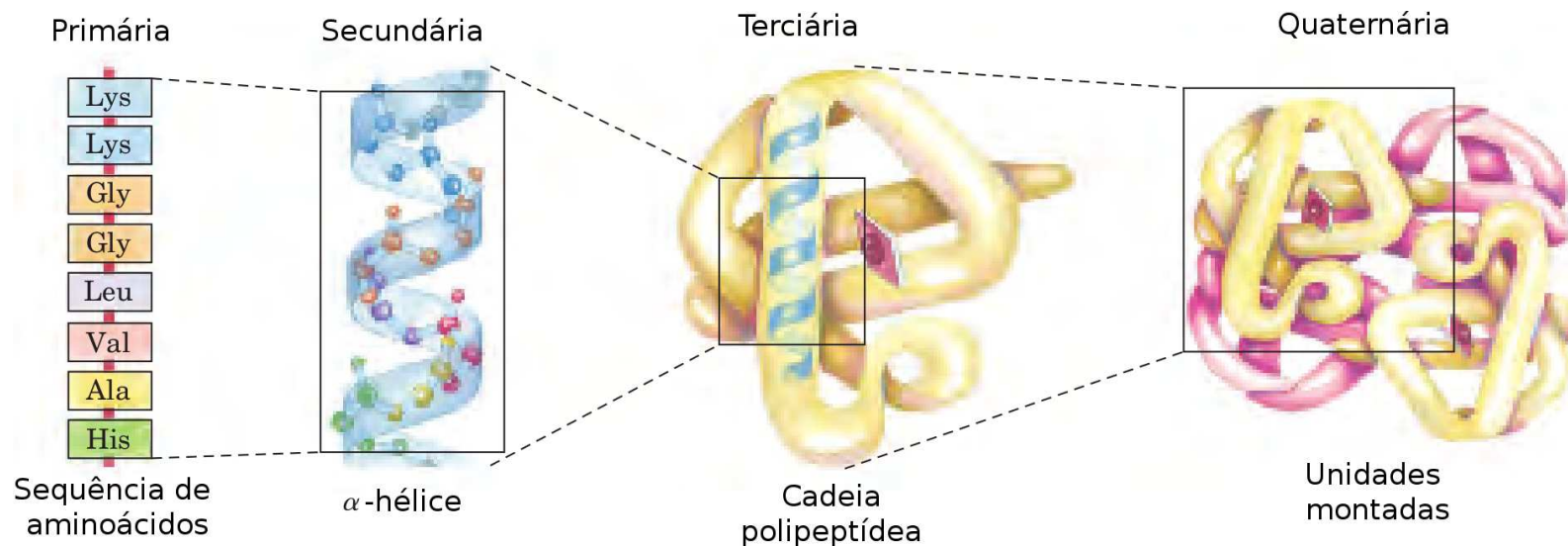


Aminoácido	R	Aminoácido	R
CYS	H ₂ C-SH	TYR	H ₂ C-  -OH
SER	H ₂ C-OH	ASP	H ₂ C-C(=O)-OH

- n tipicamente varia entre 10^2 e 10^5



Estrutura de proteínas



Perguntas fundamentais sobre proteínas:

- Como prever a estrutura 3D de uma proteína, dada a sequência de seus aminoácidos?
- Como desenhar uma proteína com uma função específica?



Prêmio Nobel de Química, 2025



The image is a blue banner for the Nobel Prize in Chemistry 2024. At the top left is a gold Nobel medal. To its right, the text reads "NOBELPRISET I KEMI 2024" and "THE NOBEL PRIZE IN CHEMISTRY 2024". At the top right is the logo of the Royal Swedish Academy of Sciences, with the text "KUNGL. VETENSKAPS-AKADEMIEN" and "THE ROYAL SWEDISH ACADEMY OF SCIENCES". Below this, three portraits of the winners are shown. Each portrait has a small vertical credit on its left: "Photo: University of Washington" for David Baker, "Photo: The Royal Society" for Demis Hassabis, and "Photo: BBVA Foundation" for John M. Jumper. Below each portrait is the winner's name and affiliation. At the bottom, two phrases are written in italics: "for computational protein design" and "for protein structure prediction".

NOBELPRISET I KEMI 2024
THE NOBEL PRIZE IN CHEMISTRY 2024

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

David Baker
University of Washington
USA

Demis Hassabis
Google DeepMind
United Kingdom

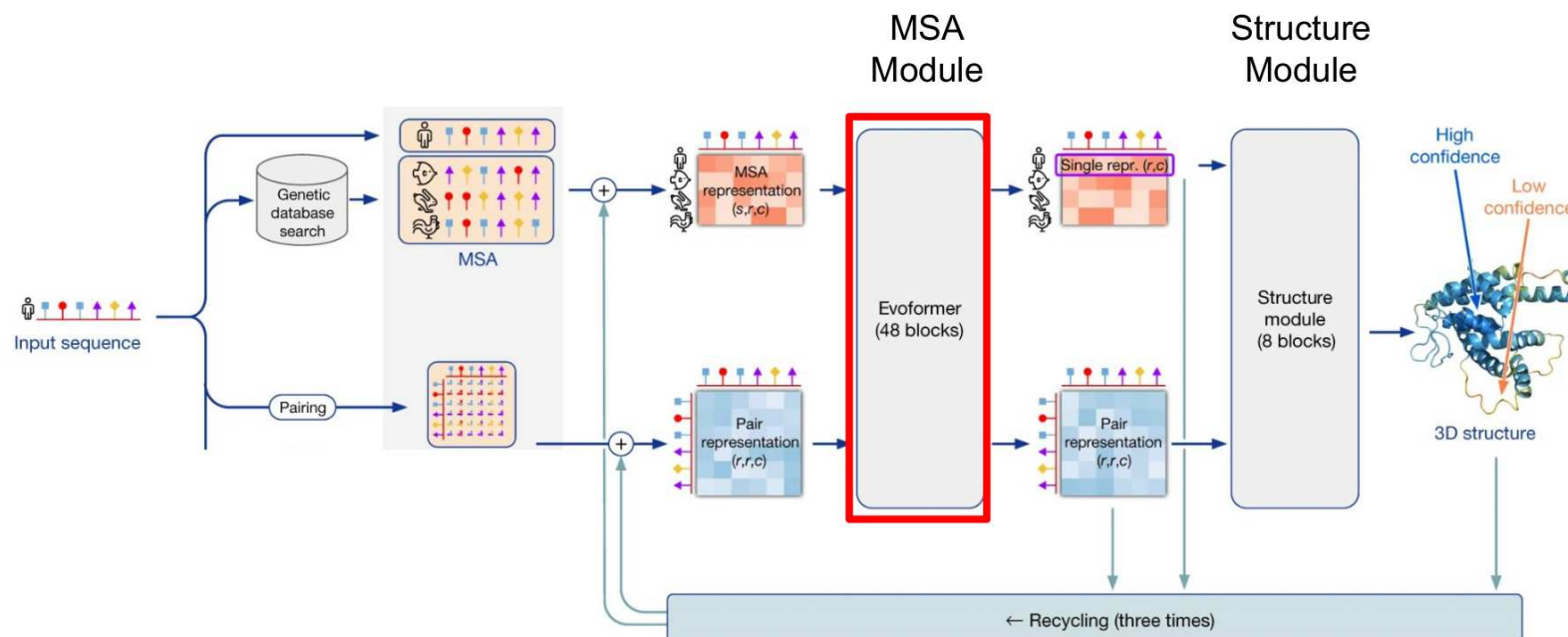
John M. Jumper
Google DeepMind
United Kingdom

"for computational protein design"

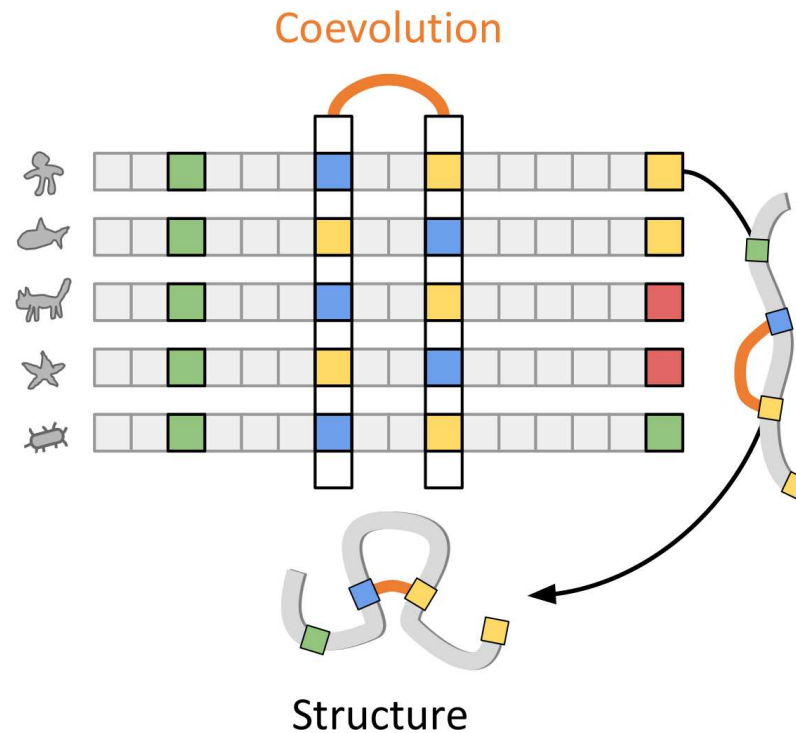
"for protein structure prediction"



Como funciona o modelo AlphaFold2?



Informação crucial: Co-evolução



- Sequências proteicas (ou genômicas) muito abundantes
→ alinhamento de múltiplas sequências (MSA)
- Usar os contatos obtidos da análise de co-evolução para guiar simulação de enovelamento!



Mais algumas aplicações com biomoléculas:

- https://github.com/biolists/folding_tools
- <https://github.com/yangkky/Machine-learning-for-proteins>
- QBQ5907: Inteligência Artificial em Biologia Estrutural
<http://gaznevada.iq.usp.br/cursos/qbq5907-inteligencia-artificial>
- <https://github.com/sokrypton/ColabFold>
- <https://github.com/sokrypton/ColabDesign>

