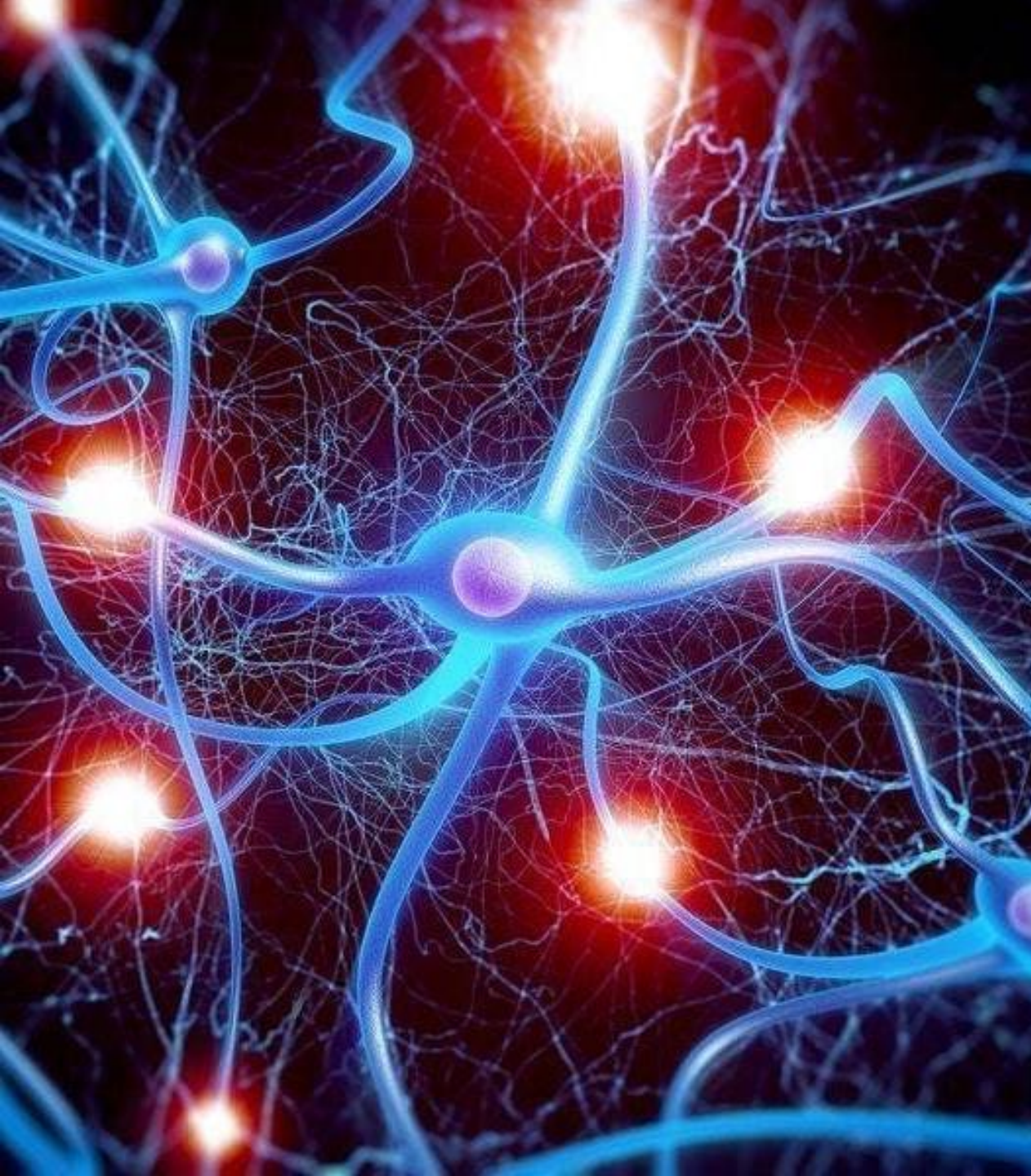




# TRANSFORMERS VÀ ATTENTION MACHANISM

VŨ TIẾN ĐẠT



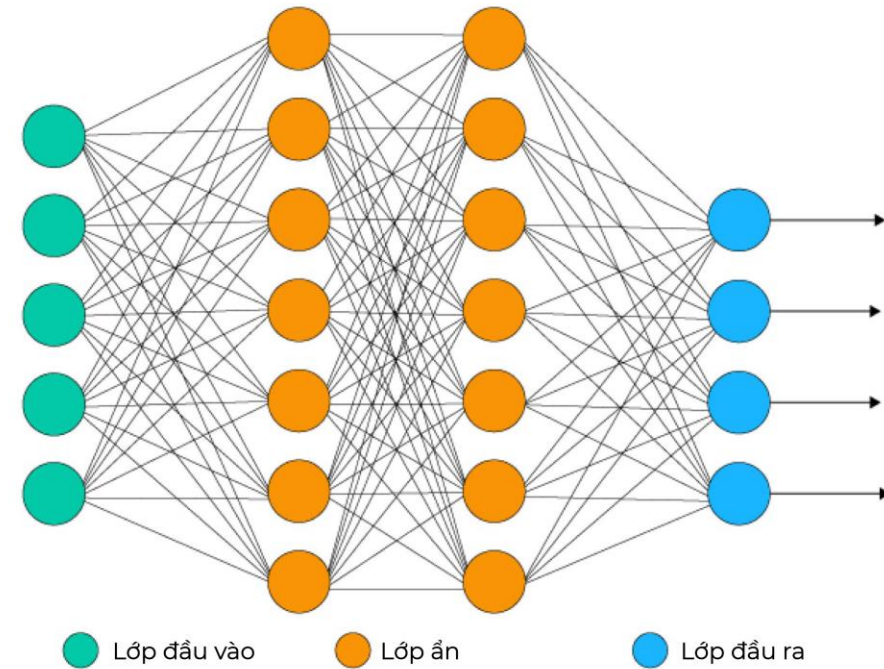


---

## TỔNG QUAN VỀ MẠNG NEURAL NHÂN TẠO

- Neural nhân tạo hoạt động dựa trên nguyên lý bắt chước các neural thần kinh trong não người. Sự bắt chước ở đây không phải là bắt chước về mặt vật lý mà chỉ là một hình thức thể hiện tự liên kết. Như trong não người, các tế bào thần kinh được liên kết với nhau bởi các khớp nối thần kinh, và nó là một thực thể vật lý. Chính vì sự liên kết như vậy tạo nên một mạng lưới thần kinh lớn, neural này liên kết với neural kia, một khi có xung điện được truyền đi thì tín hiệu của 1 neural cứ như vậy sẽ được lan truyền qua các lớp sau theo cấu trúc mạng lưới.
-

- ❖ Trong y học đã chứng minh nhờ sự liên kết như vậy mà mới tạo nên tư duy của loài người. Và đó chính là cơ sở nền tảng đầu tiên cho việc hình thành ra Neural nhân tạo.
- ❖ Như trong hình có thể thấy các tế bào thần kinh nhân tạo được liên kết với nhau một tế bào của lớp đằng trước sẽ liên kết với toàn bộ tế bào neural ở lớp đằng sau và đó hoàn toàn là giống như cấu trúc của mạng thần kinh tự nhiên
- ❖ Hình ảnh về mạng thần kinh tự nhiên như trong não người là tồn tại thật sự, tuy nhiên mạng thần kinh nhân tạo được vẽ như trong ảnh chỉ là một cách thể hiện ngôn ngữ hình ảnh của con người trước một vấn đề quá trừu tượng như AI. Về bản chất mọi thứ đều nằm ở từ liên kết.



---

Liên kết chính là từ khóa chỉ mạng gỡ rối mọi thứ trước khi học AI. Bản thân tôi đã từng gặp nhiều người không học được vấn đề này là vì không nắm được những vấn đề chủ chốt mà chỉ đi đâm đầu cho nhanh ra thành quả bằng cách sử dụng những mã nguồn có sẵn. Dẫn đến câu chuyện họ không ứng dụng được những lý thuyết và tư duy hay ho của lĩnh vực này cho mục đích cuộc sống. Vì vậy liên kết chính là chìa khóa để chinh phục được AI.

Các tế bào thần kinh tự nhân tạo thực chất nếu viết trên giấy thì nó chỉ là những hàm số (Ánh xạ) và chúng liên kết với nhau thông qua liên kết logic (ẩn số) ( Bỏ qua việc suy xét liên kết về mạch điện tử ở CPU).

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases}$$



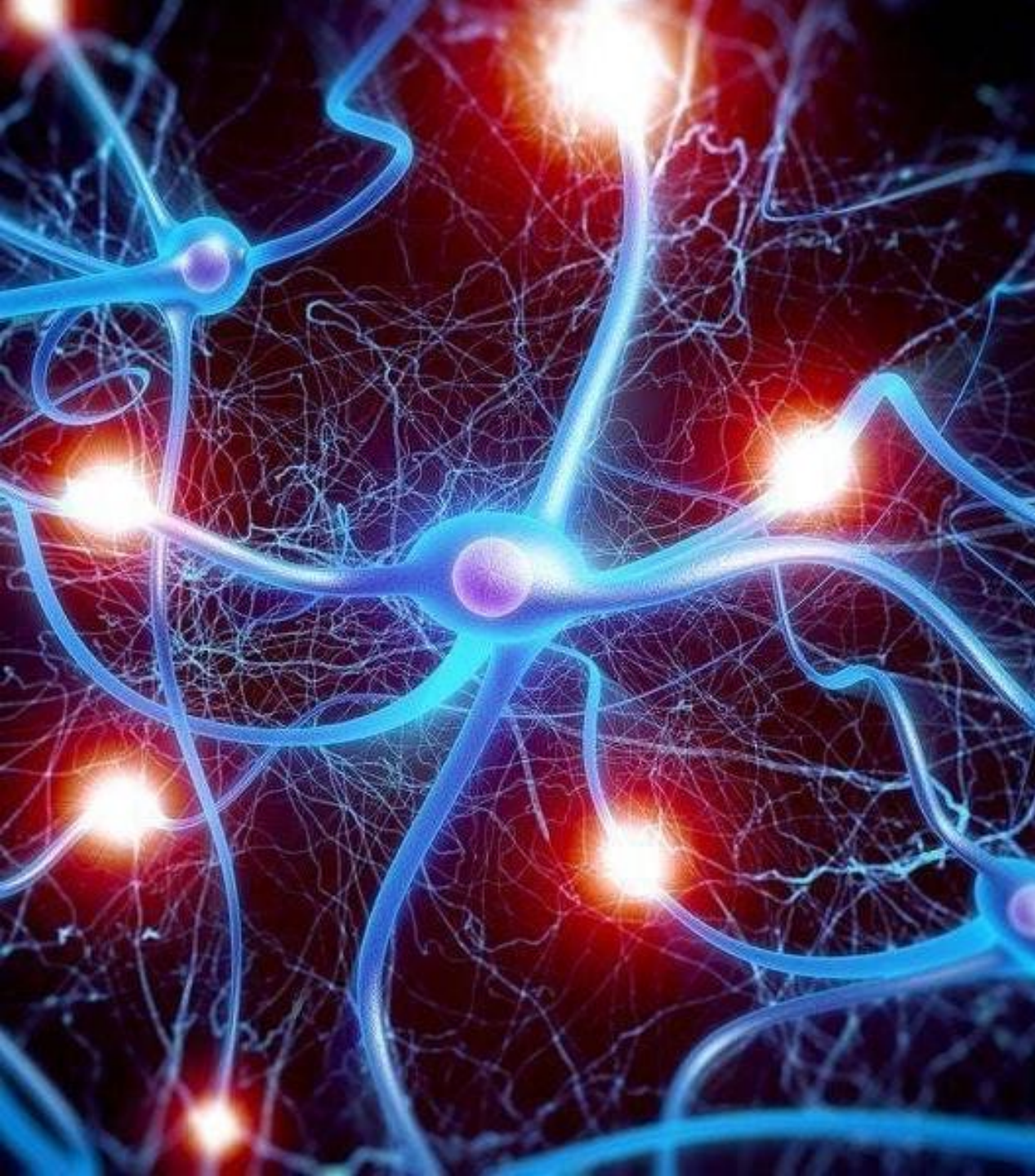
---

Như mọi người cũng có thể thấy giả sử như hai hệ phương trình trên là hàm hai biến thì rõ ràng phương trình 1 đã có sự liên kết với phương trình 2 và ngược lại vậy cho nên đây chính là sự liên kết giữa hai tế bào thần kinh và nếu nhiều tế bào hơn nữa thì có thể mở rộng hệ phương trình ( Số chiều ) ra. Như vậy đó chính là liên kết của mạng thần kinh. Còn khi muốn mạng thông minh hơn, chính xác hơn thì ta sẽ phải làm cho hệ càng phi tuyến càng tốt. Khi đó ta xem như hệ phương trình là hệ phương trình phi tuyến

$$\begin{cases} x + y + xy = 1 \\ x^2 + y^2 + 3(x + y) = 28 \end{cases}$$

Nguyên nhân vì sao lại như vậy thì qua phần sau ta sẽ làm rõ hơn, nhưng đại loại là làm cho não có nhiều nếp nhăn hơn. Kiểu là như vậy (cách để hình dung nhưng sai về ý đồ toán học)

---



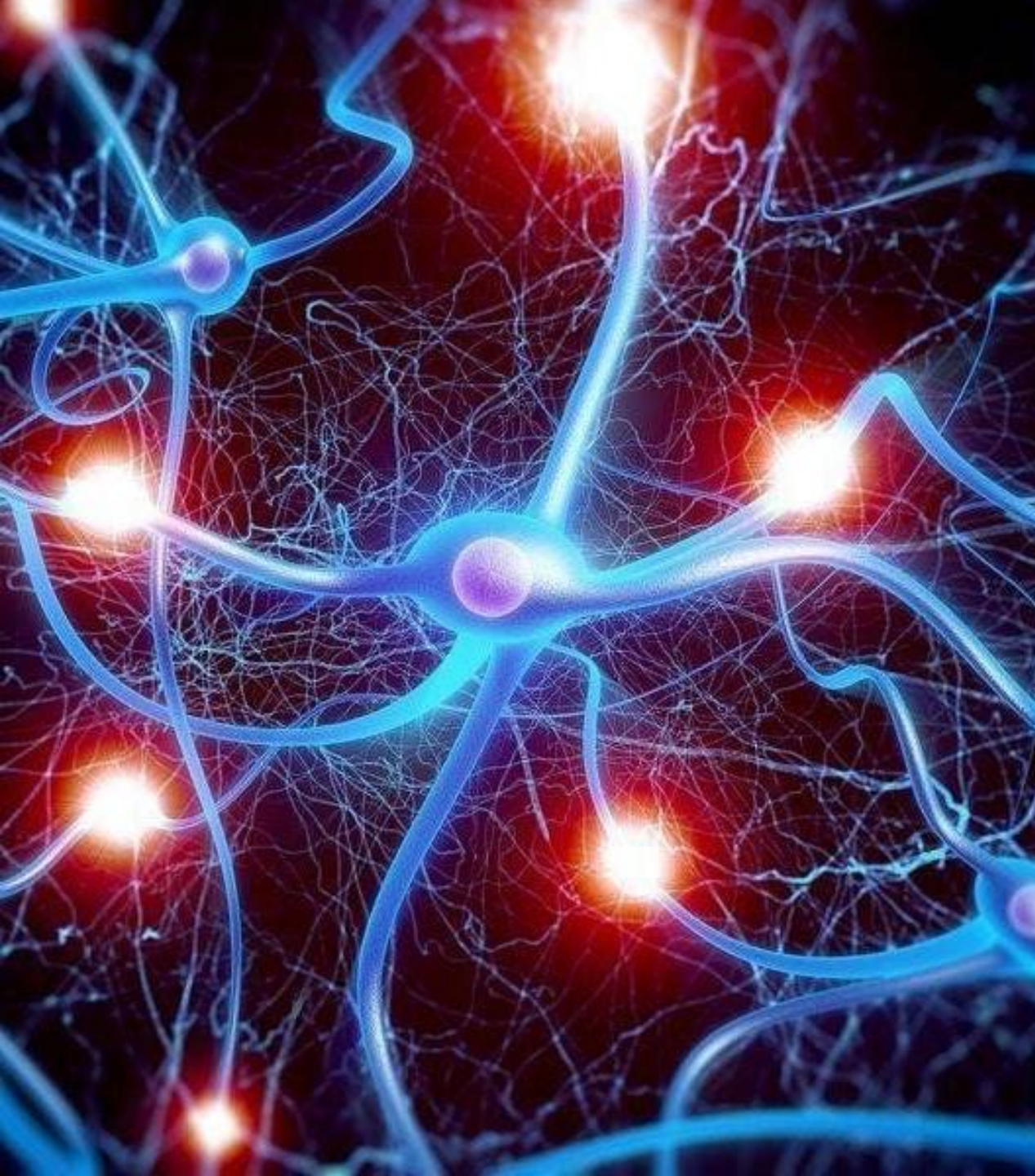
---

## TỔNG QUAN VỀ TẾ BÀO THẦN KINH (PERCEPTRON)

Tế bào thần kinh là đơn vị cơ bản để cấu tạo nên mạng thần kinh. Trong mạng neural nhân tạo khi viết ra giấy thì nó chính là một hàm số (trong code cũng vậy). Hàm số thì có nhiều loại ví dụ như hàm tuyến tính, phi tuyến, liên tục trơn, không liên tục convex,... Hàm số ta sử dụng trong neural là hàm phi tuyến. Vậy để hiểu được cách tư duy của mạng như thế nào thì ta phải hiểu cách mà một tế bào tư duy như thế nào để giải quyết được vấn đề, từ đó sẽ mở ra một chân trời khám phá phía sau mà không bị lẫn lộn.

---





---

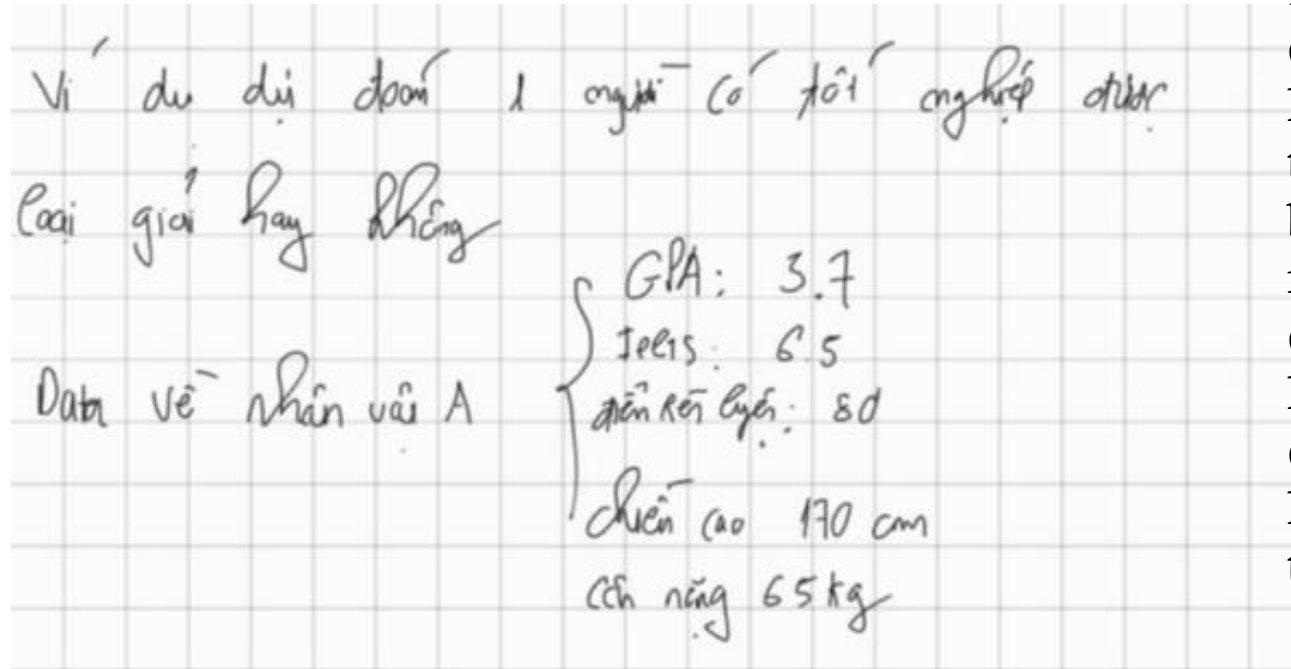
## TỔNG QUAN VỀ TẾ BÀO THẦN KINH (PERCEPTRON)

Theo kinh nghiệm cá nhân của tôi thì việc học AI bao gồm 3 bước như sau

- 1) Hình dung được liên kết tức là hiểu được cách thức vận hành
  - 2) Hình dung được toán học áp dụng như thế nào để khiến mạng tư duy được
  - 3) Hình dung được cách chuyển từ tư duy nhị phân sang tư duy cảm xúc
  - 4) Áp dụng toán học một cách linh hoạt để tinh chỉnh mạng
  - 5) Coding và thử nghiệm sẽ xác thực lại được toàn bộ lý thuyết mà mình đang biết
-

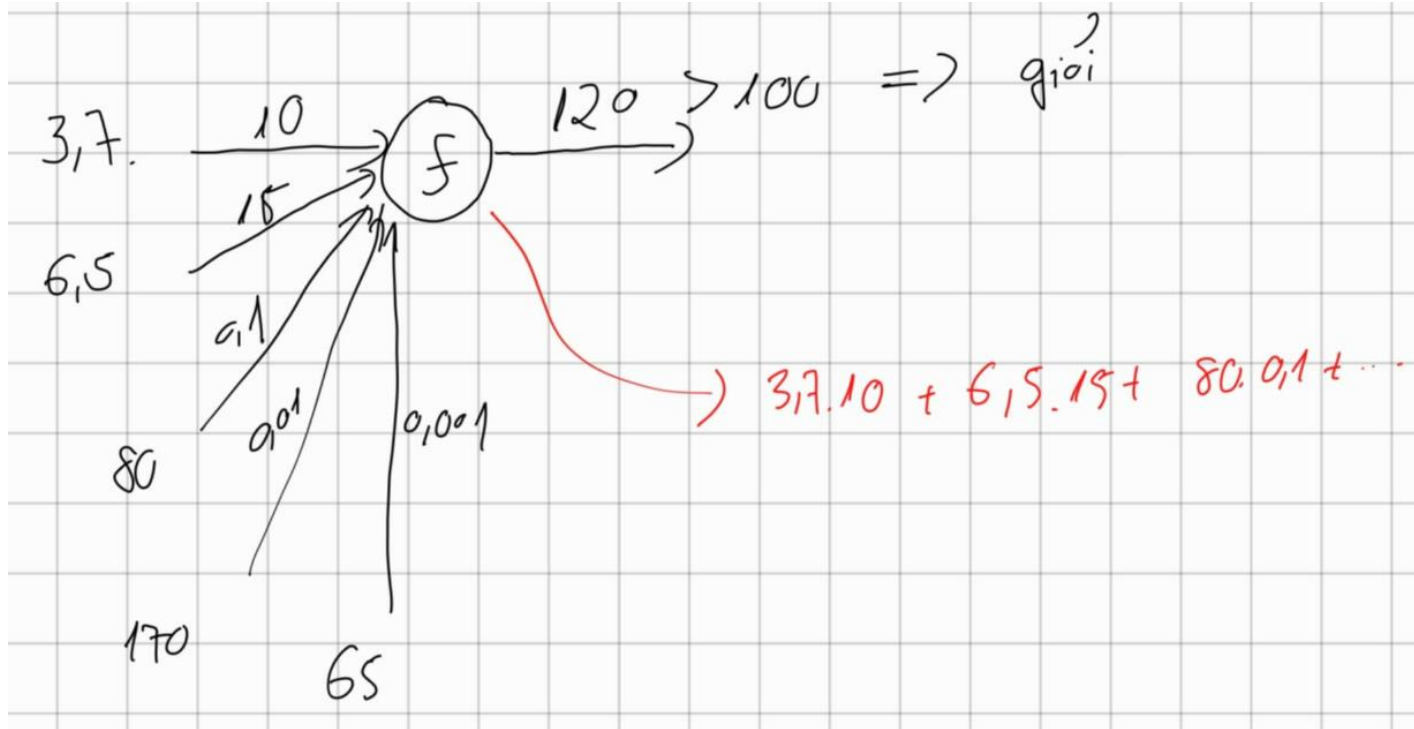
---

## TẾ BÀO THẦN KINH NHÂN TẠO: PERCEPTRON

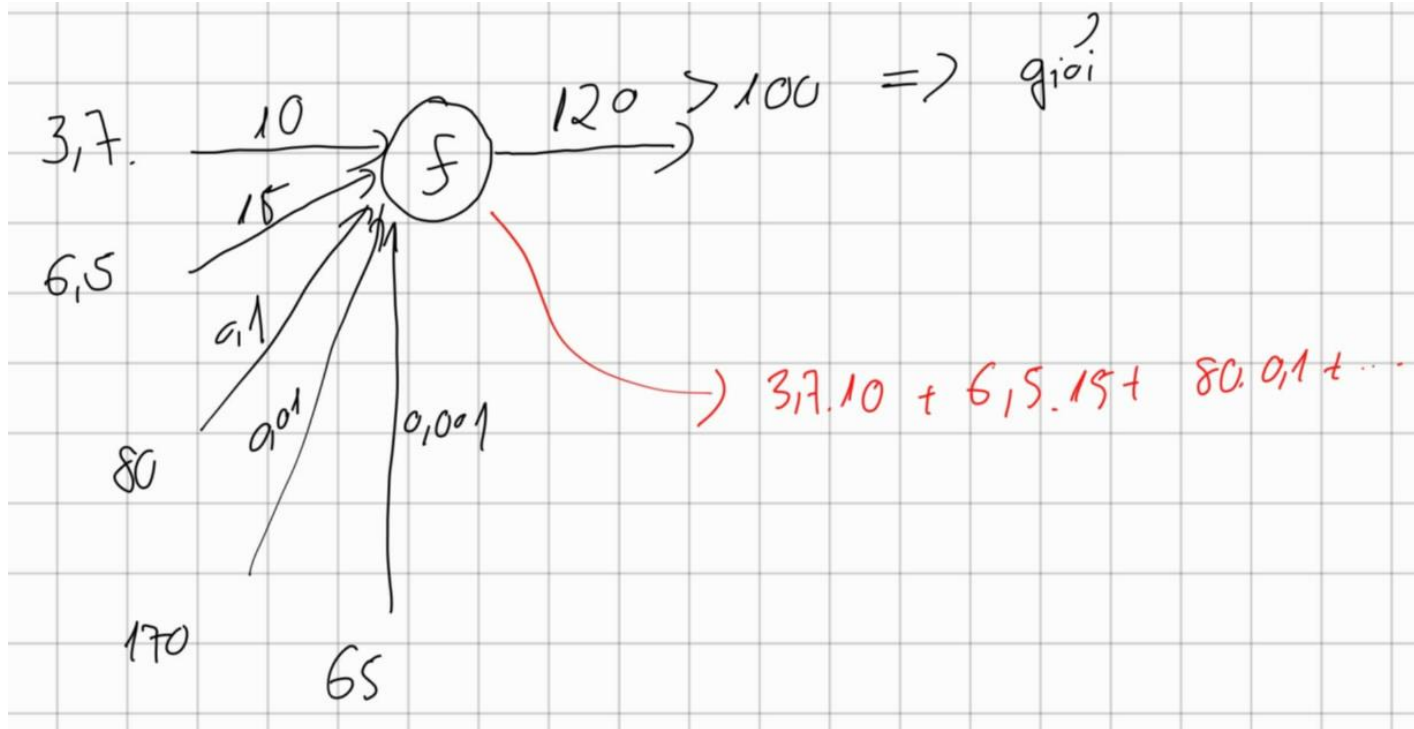


Đối với ví dụ trên ta sẽ sử dụng mạng neural để dự đoán người này có được đạt danh hiệu sinh viên giỏi hay không. Trong AI quá trình huấn luyện mạng chính là đi tìm ra các bộ trọng số phù hợp. Giả sử như mạng neural tôi đã huấn luyện xong và đã tìm được bộ trọng số phù hợp rồi thì tôi sẽ tiến hành dự đoán như sau. Với giả định tròn đoạn code tôi sẽ code là nếu như hàm  $f$  nhỏ hơn 100 thì là không giỏi, còn lớn hơn 100 thì mới giỏi.



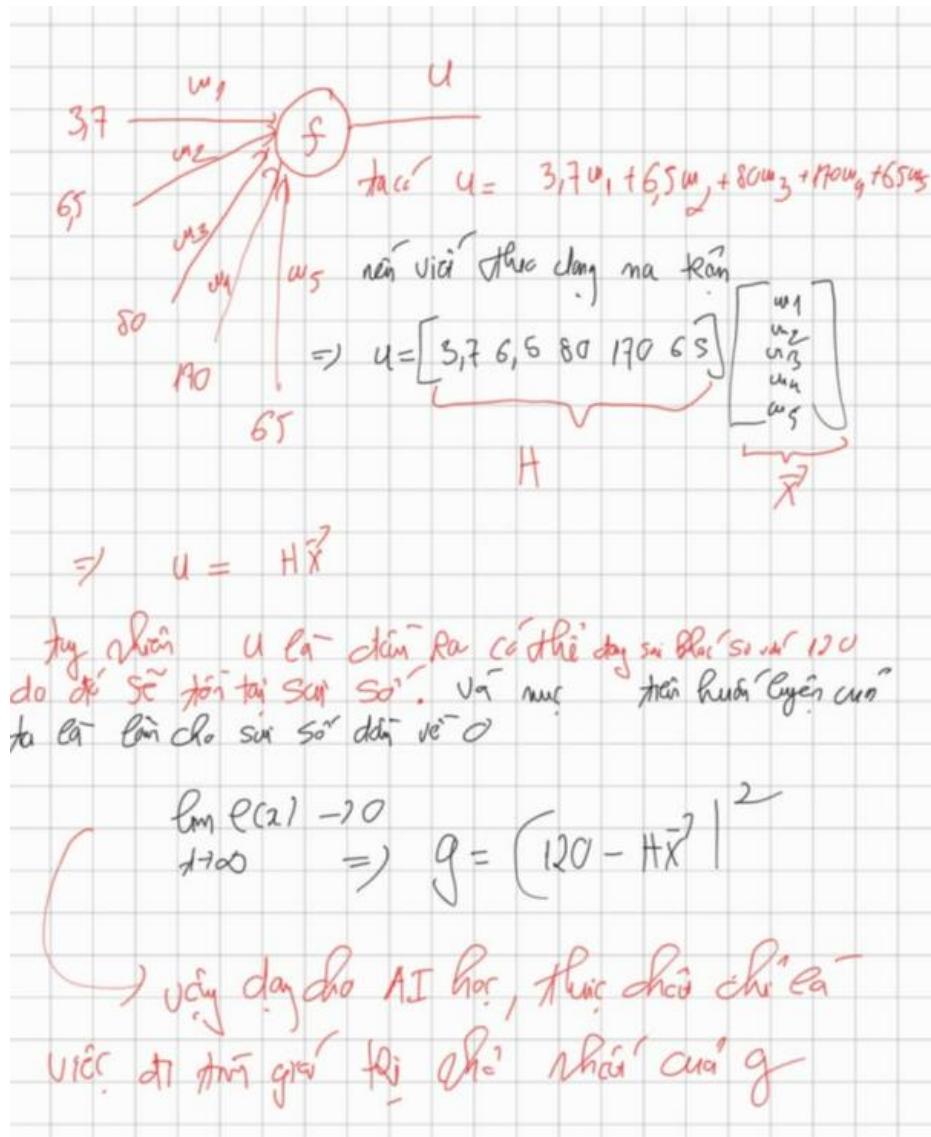


Và đây chính là cách mà một neural network sẽ so sánh với giá trị ngưỡng để ra quyết định. Và AI chỉ có một cách tư duy duy nhất đó chính là so sánh với một ngưỡng như vậy thôi. Và vấn đề trở nên hay hơn là chúng ta sẽ có những thủ thuật toán học để từ việc so sánh với một ngưỡng như thế mà có thể chuyển qua những thứ khác như là trò chuyện với con người hoặc này kia khác nó. Cái đó tôi sẽ giải thích sau, sau khi người đọc hiểu được những tư duy cơ bản này. Ở đây tôi gọi là tư duy cơ bản chứ không phải kiến thức cơ bản, bởi vì với tôi khi tiếp cận với những vấn đề này nó thiên về tư duy hơn



Cái cách mà mạng thần kinh tư duy cũng đơn giản, trên những cạnh liên kết ta gắn kèm những trọng số và đối với những thông tin quan trọng cho việc xác định người đó có giỏi hay không như GPA, hay ielts thì sẽ được đánh trọng số cao, còn những thông tin không liên quan như điểm rèn luyện, chiều cao cân nặng,... Không liên quan thì sẽ được gắn trọng số thấp hơn thể hiện là không quan trọng. Trong quá trình huấn luyện mạng thì sẽ tự xác định được điều đó.





- Như vậy bài toán ví dụ trên tôi đang giải theo kiểu giải xuôi tức là tôi đã có sẵn trọng số rồi và chỉ việc lồi ra để tính và ra được kết quả thì đó gọi là bài toán dự đoán. Còn đặt giả định hiện tại tôi có dữ liệu là người đó với những thông tin như vậy và người đó còn được học sinh giỏi thì liệu rằng với người khác cũng có những thông tin như vậy thì liệu có được học sinh giỏi không. Từ đó mới dẫn tới câu chuyện tôi sẽ đi xác định một bộ trọng số dựa trên thông tin đã có và để đi sử dụng cho việc dự đoán ở chỗ khác.