
U22 PROCON

AIの思考が「のぞける」「さわれる」教育アプリ

LLMView

30 November, 2025

AIの思考が「のぞける」「さわれる」教育アプリ

LLMView

CONTENTS 01

AIの思考を「のぞく」

「AI」 それ自体をおもしろく

NOW

LLMView

高品質

期待通りの結果

過程よりも結果

素早さ

体験よりも実用性

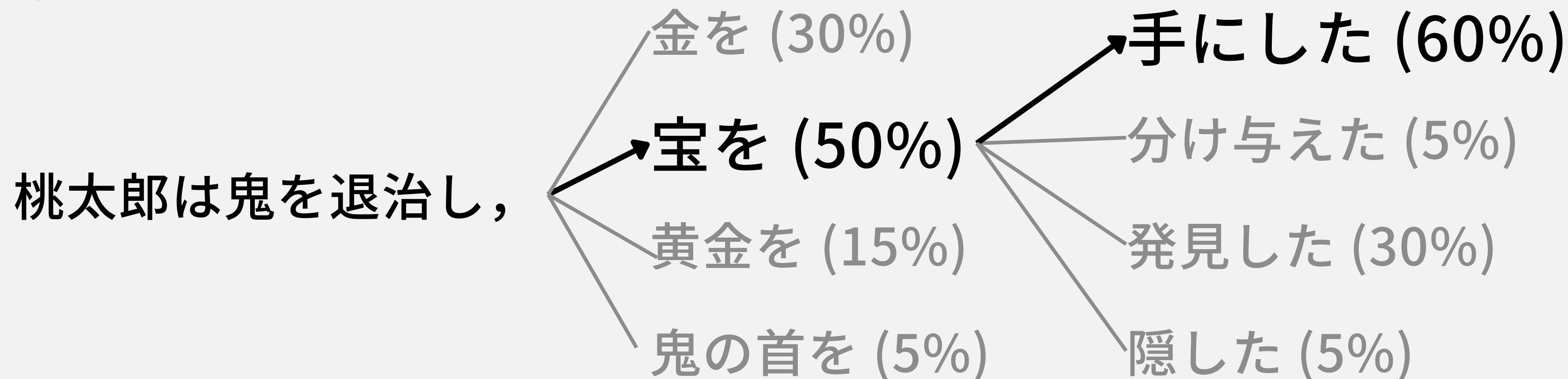
AIの「**迷い**」をコンテンツに

思考の「**分岐**」に直接触る

「使う」よりも「**いじる**」楽しさ

「AI」の思考をのぞく

Q.桃太郎は何をした？



AIの思考が「のぞける」「さわれる」教育アプリ

LLMView

CONTENTS 02

AIの思考を「さわる」

CONTENTS 02 AIの思考を「さわる」

LLMView

ナビゲーション

チャット履歴

18と23の積は？
1メッセージ・2025/11/24
13:01:13

18と23の積は？
1メッセージ・2025/11/24
12:56:11

18と23の積は？
1メッセージ・2025/11/24
11:51:26

47+53は？
1メッセージ・2025/11/24
11:47:53

何をしてるの？
1メッセージ・2025/11/24
11:35:24

設定

LLMView AI可視化ツール

+ 新規チャット

18と23の積は？
22:58

18と23の積は、 $18 \times 23 = 414$ です
22:58
[クリックしてツリーを表示](#)

メッセージを入力... 送信

馴染みのある
チャット形式のUI
扱いやすく、
「いつものAI」の
裏側をのぞくような
体験を演出

CONTENTS 02 AIの思考を「さわる」



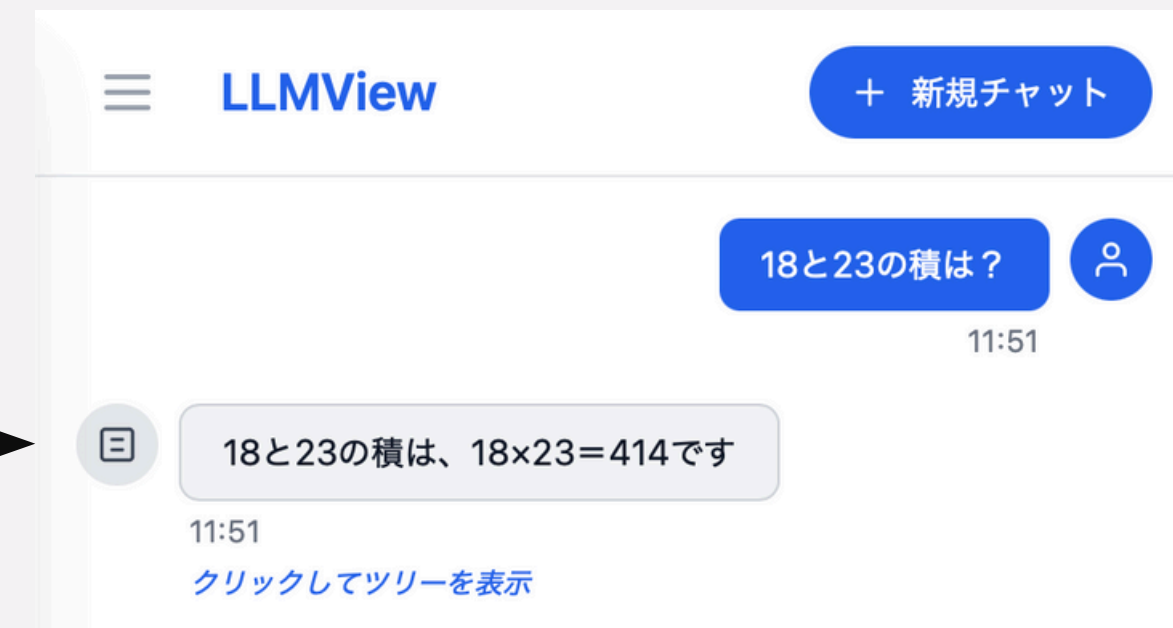
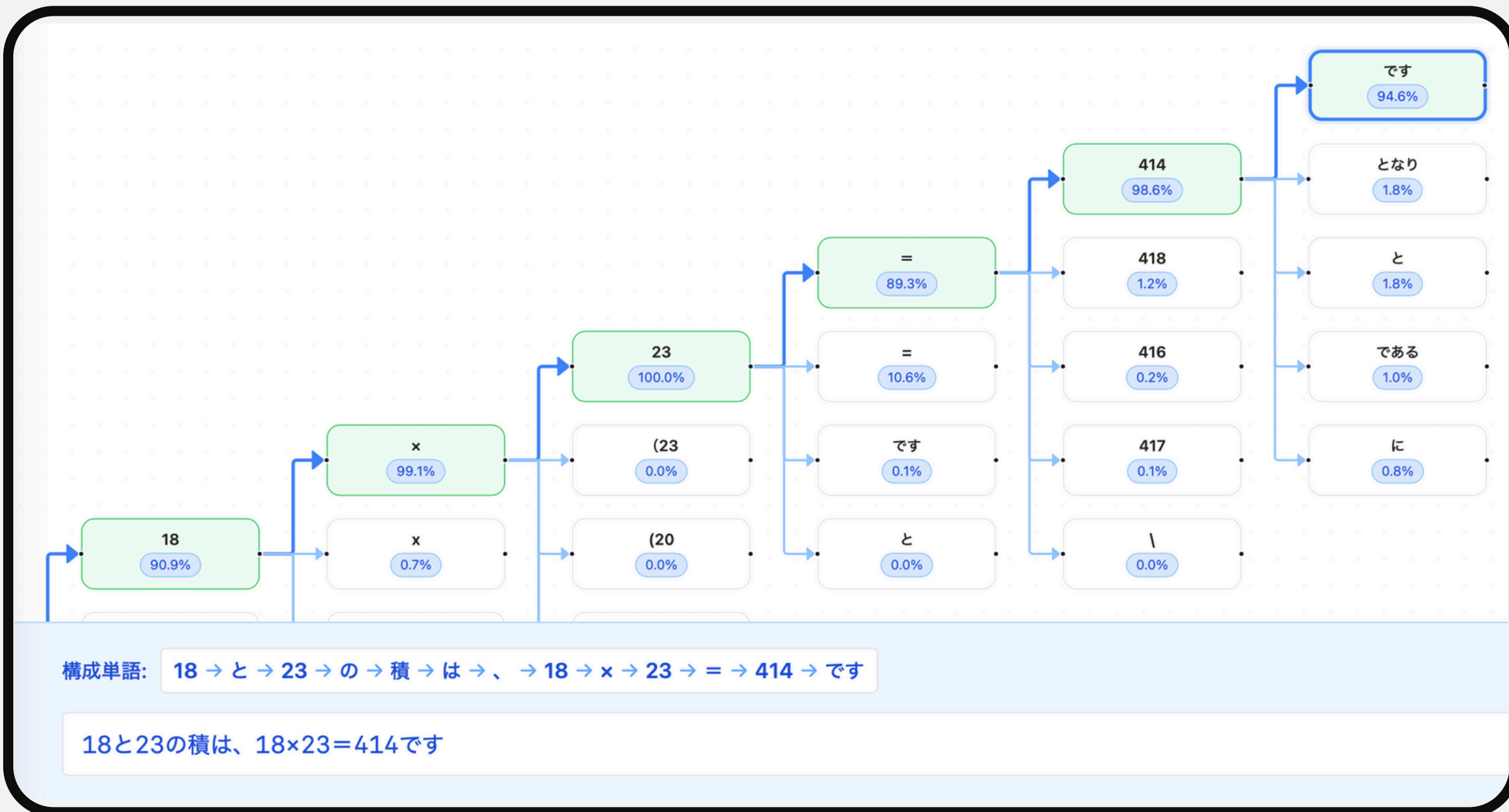
「オリジナル」の単語を
ツリーに追加

次の単語候補をAIで生成

「次の候補」が生成された
ノードは緑色で表示

選択したノードまでの
文章をチェック可能

CONTENTS 02 AIの思考を「さわる」



おなじみのチャットから裏側をのぞくような「体験」



「IF」の世界線を切り拓く



AIの中身を直感的に「学ぶ」

CONTENTS 02 AIの思考を「さわる」

01 構造の理解

「AI」は魔法ではなく
確率に基づく「論理」

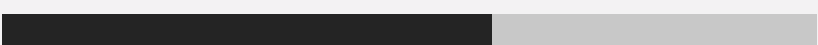
解像度: 30%



02 IFの探求

95%の「正解」に隠れた、
5%の「可能性」を選ぶ

解像度: 60%



03 自由自在

「使う」から「操る」へ
自分で新たな分岐を作る。

解像度: 100%



AIの思考が「のぞける」「さわれる」教育アプリ

LLMView

CONTENTS 03

AIの思考を「動かす」

実装におけるの壁

人間とAIの言語の差

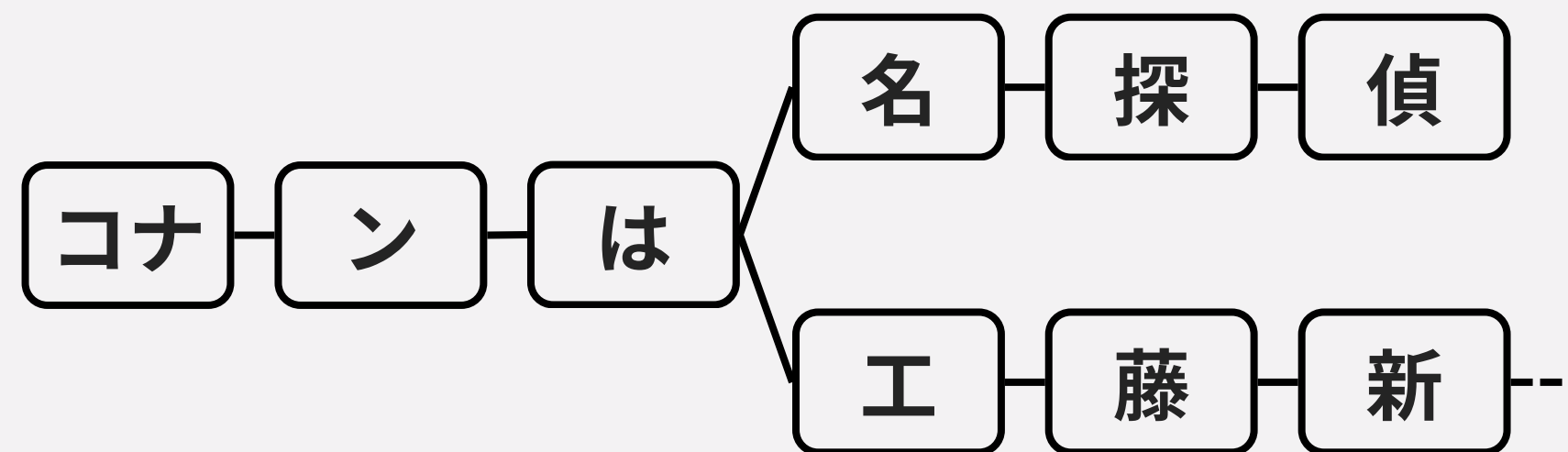
人間：コナン / は / 名探偵

と単語の区切りで把握する

AI：コナ / ン / は / 名 / 探 / 偵

と1、2文字の塊で生成する

そのまま実装すると...



➡ 見にくい・読みにくい

アルゴリズムによる解決

ストリーム単語決定アルゴリズム

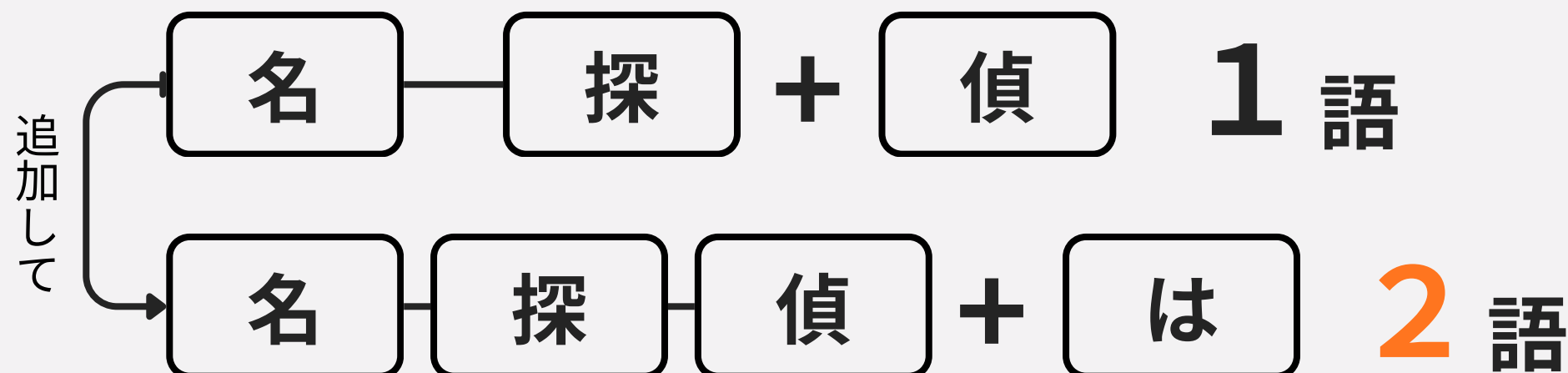
課題

AIの認識するToken
を人が読める単語に

制約

「先読み」ができない
ストリームデータ

単語決定アルゴリズム

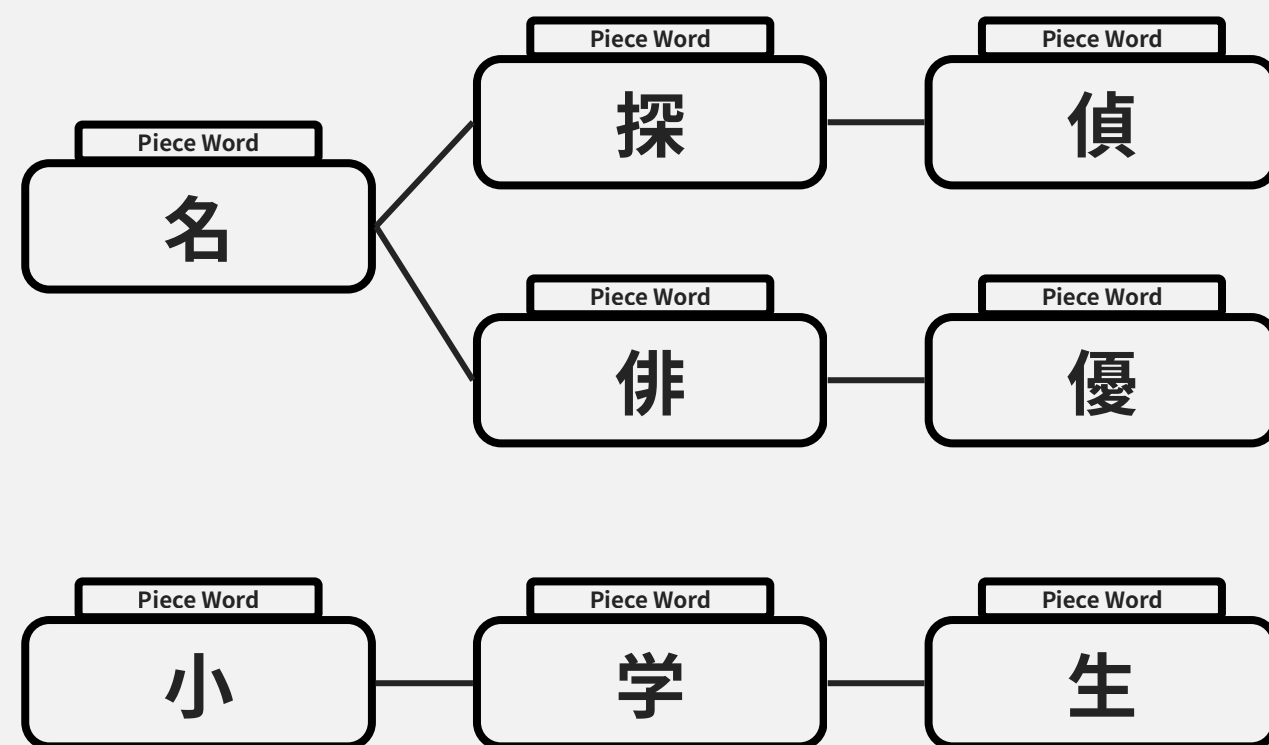


➔ 名探偵 を単語として確定

✓ 1文字ずつのTokenから 1単語に

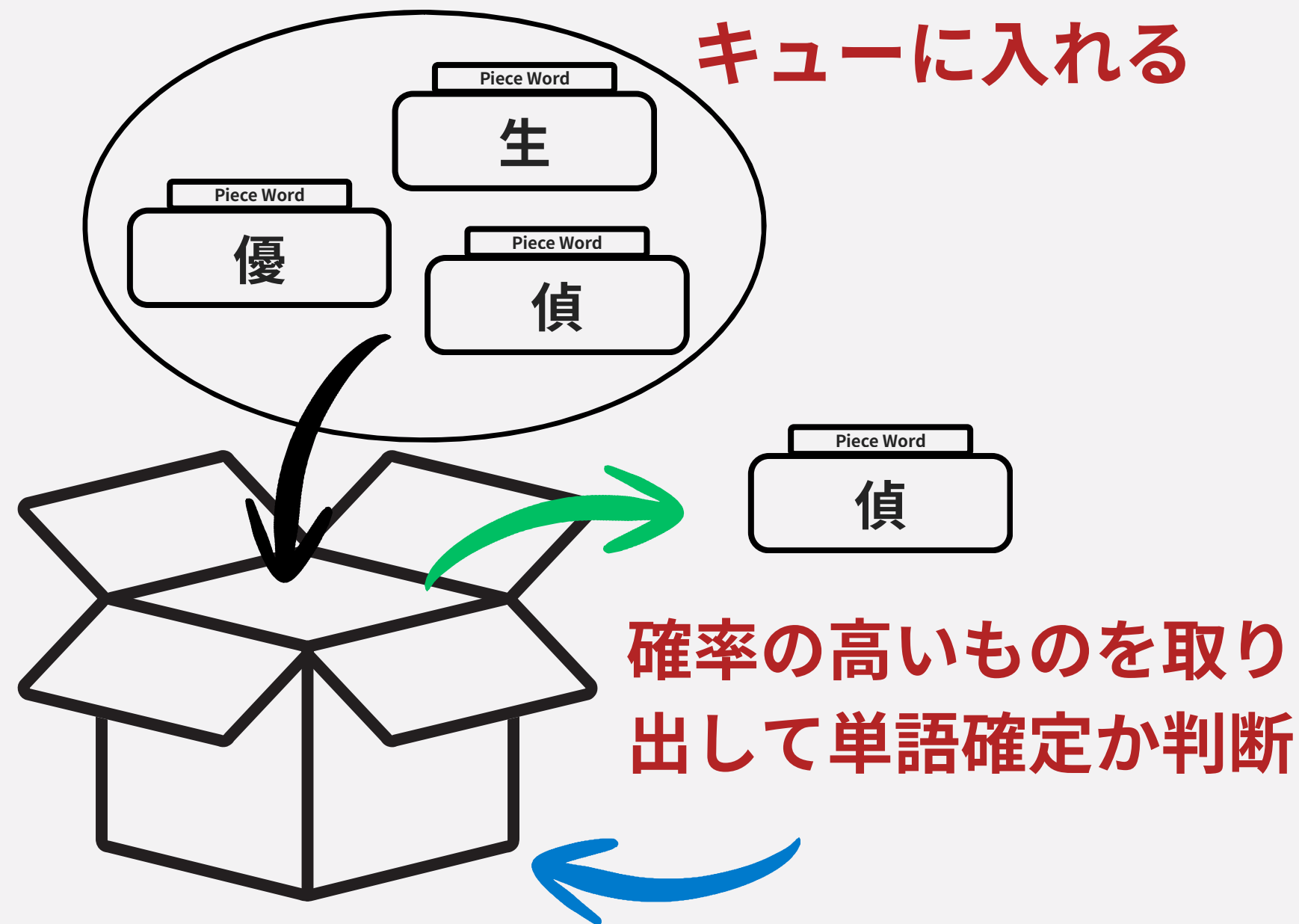
【データ構造】

「コナンは」の続き



→候補をツリーの形式で管理

末端を優先度付き
キューに入れる



確率の高いものを取り
出して単語確定か判断

正しい分割を判断する仕組み

ストリーム単語決定アルゴリズム

課題

「ありが(とう)」を
「あり / が」と間違える

制約

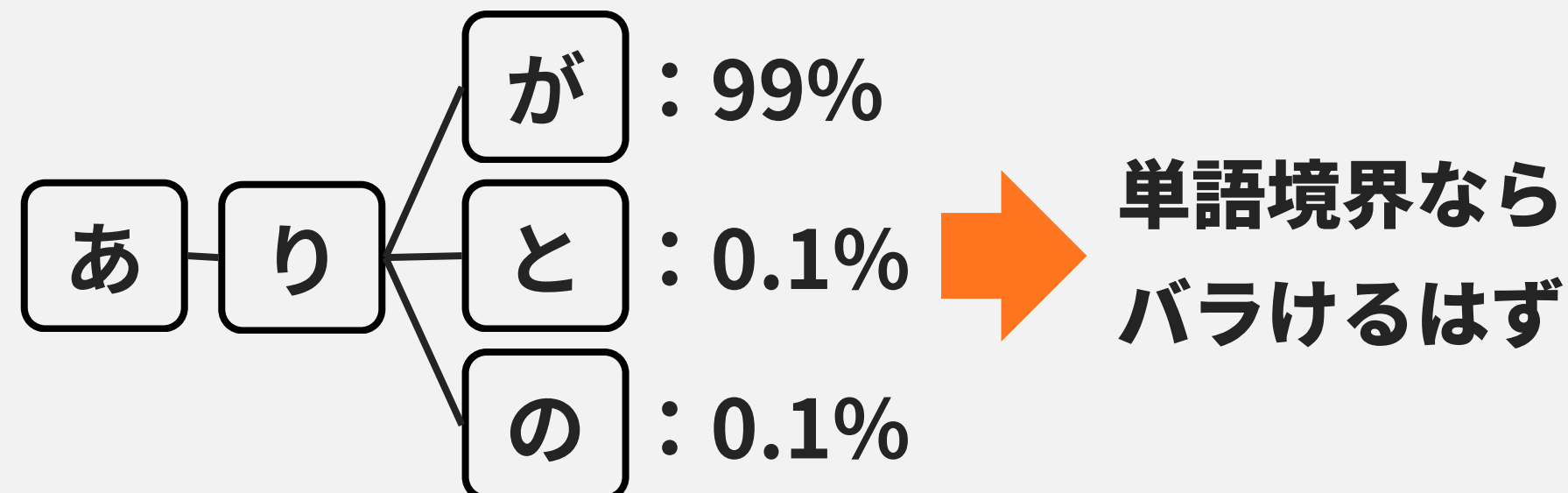
「先読み」ができない
ストリームデータ

単語決定アルゴリズム

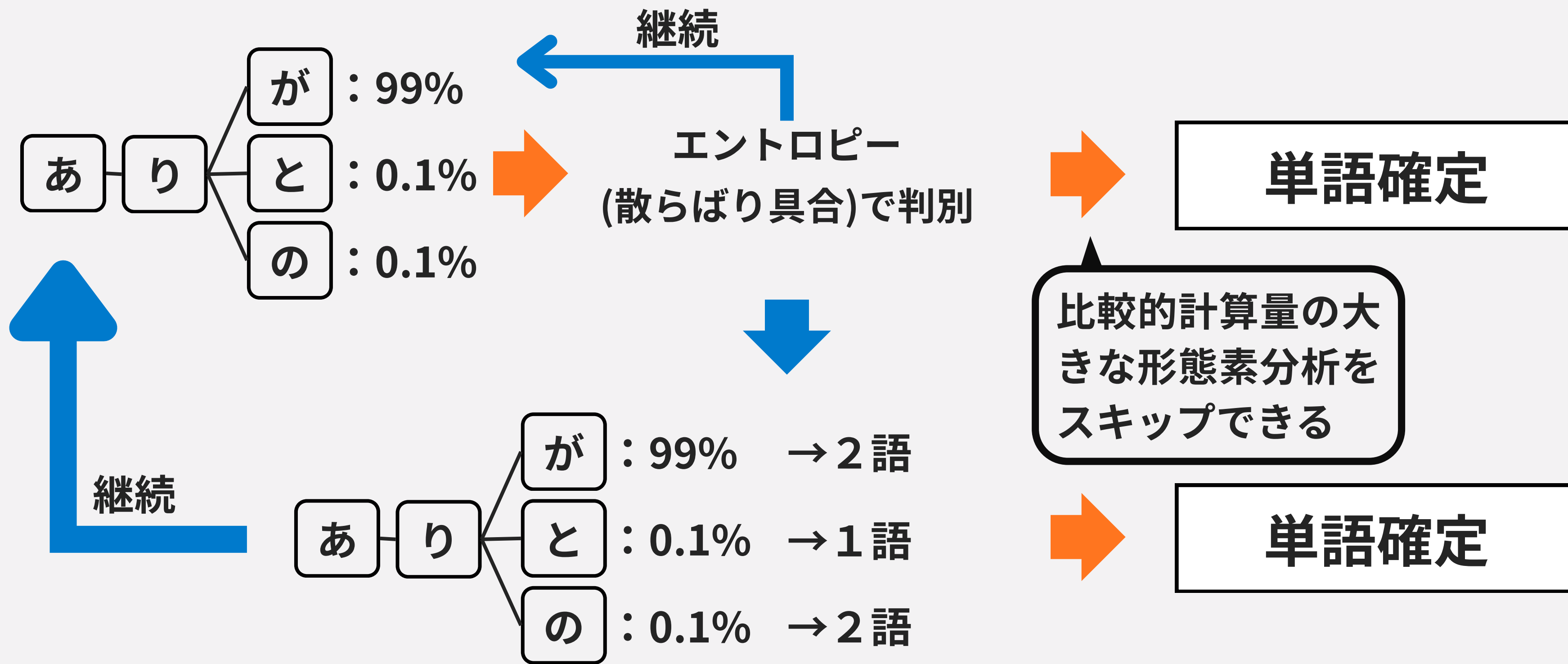
あ — り + が 2語？

制約 を利用したアルゴリズムで解決

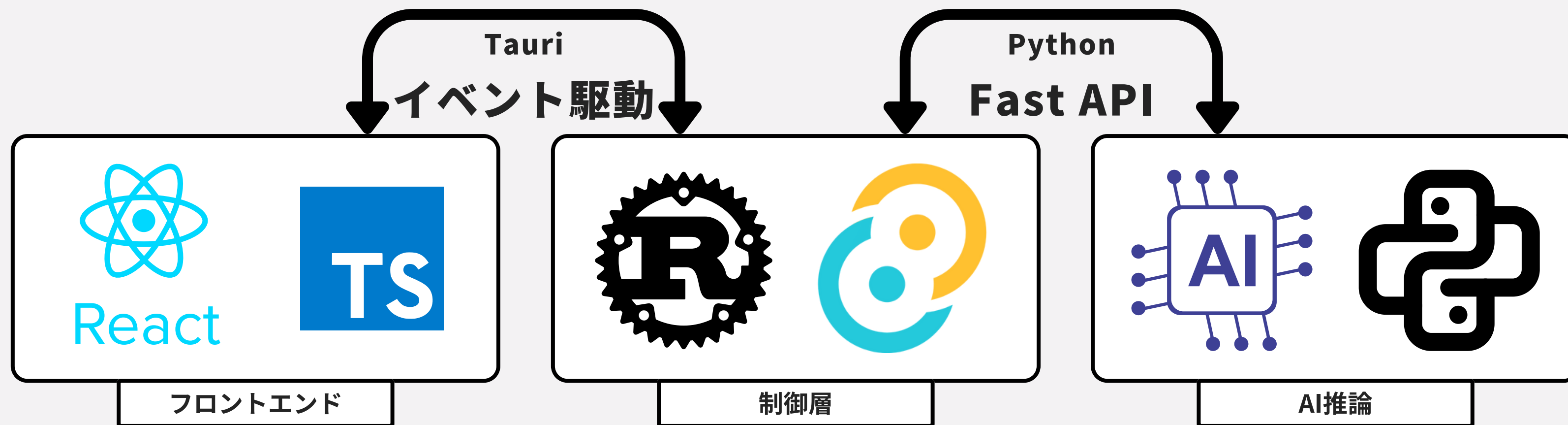
➡ 「候補と確率」から単語決定を予測



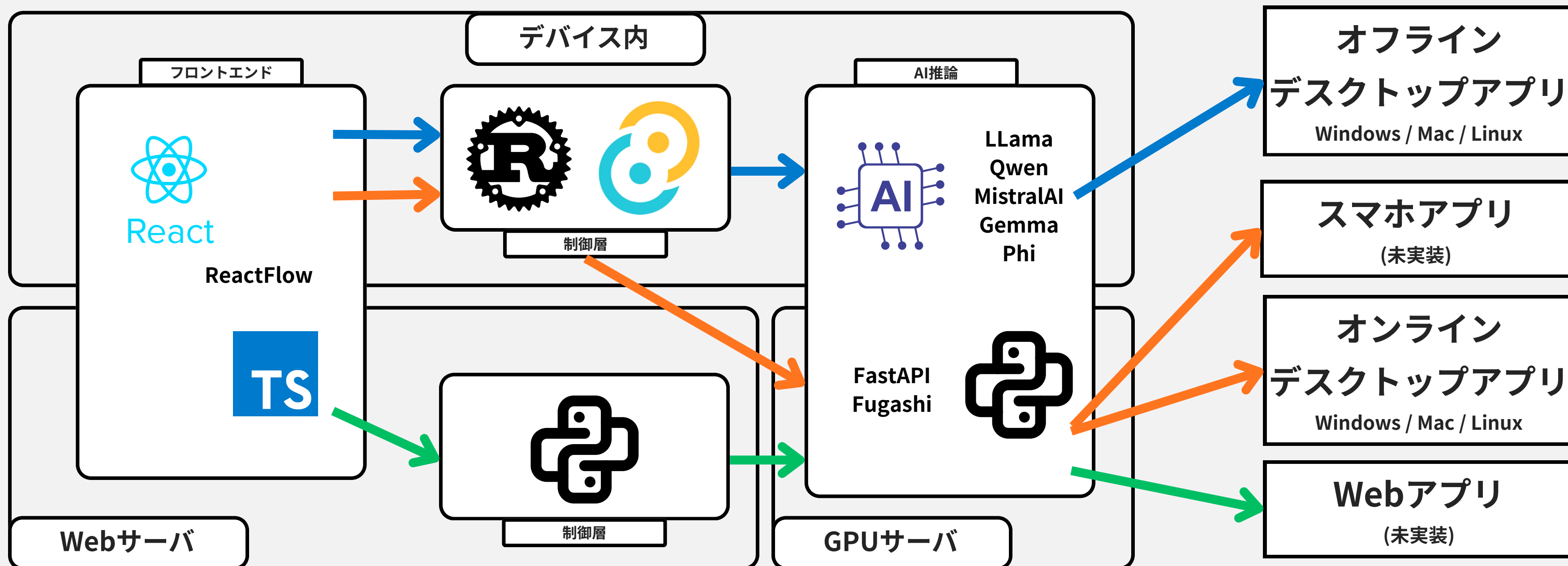
CONTENTS 02 AIの思考を「動かす」



UI・制御・推論の「完全分離」



構成を「スイッチ」する柔軟性



構成を「スイッチ」する柔軟性

01 実装スピード フロントエンド & AI推論層をプラットフォーム間で共通化し、
迅速なサービス展開を可能に。

02 スケール可能 AI層が独立しており、ユーザー増に合わせた
GPUサーバーの増設でサービス規模を柔軟に調整可能

03 デプロイ戦略 個人利用は「ローカル」、大規模展開は「クラウド」
ユーザに合った使用方法を提供

U22 PROCON

AIの思考が「のぞける」「さわれる」教育アプリ

LLMView

30 November, 2025

【プロンプト】

Llama 3 / 3.1 / 3.2

```
<|start_header_id|>system<|end_header_id|>
```

```
    You are a helpful assistant.<|eot_id|>
```

```
<|start_header_id|>user<|end_header_id|>
```

```
    Hello world!
```

```
<|eot_id|>
```

```
<|start_header_id|>assistant<|end_header_id|>
```

```
    Hello I am AI.
```

```
<|eot_id|>
```

Phi-3 / 3.5 (Mini, Medium, Vision)

```
<|system|> You are a helpful assistant.<|end|>
```

```
<|user|> Hello world!<|end|>
```

```
<|assistant|> I am AI. <|end|>
```

Qwen 2 / 2.5

```
<|im_start|>system
```

```
    You are a helpful assistant.
```

```
<|im_end|>
```

```
<|im_start|>user
```

```
    Hello world!
```

```
<|im_end|>
```

```
<|im_start|>assistant
```

```
    I am AI.
```

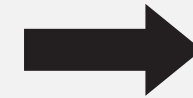
```
<|im_end|>
```



「Assistantの開始タグ（ヘッダー）」を付け足した状態で入力を渡すと、
モデルがその続き（返答の中身）を生成し始める。

【Logprobs(次トークンの対数確率リスト)】

Tokenの一覧	Logprobs
Token 1	-4.6
Token 2	-6.9
....	...
....	...
Token 1万	-0.1



ソートして上位のToken
を得る (TopK)



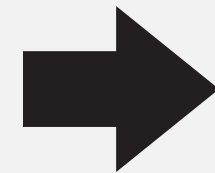
現状は5個にしているが、10個や20個、3個なども対応可能(多くなるほど分岐が多くなり、時間がかかるので、無制限に多くはできないが)

【KVCache(生成途中のCacheの再利用)】

私の名前はコナン・ドイルです。趣味は....

↓ 3秒

サッカーと野球です。



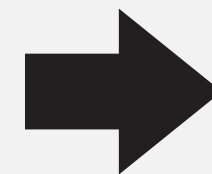
私の名前はコナン・ドイルです。趣味はサッカーです。休みの日は

私の名前はコナン・ドイルです。趣味はサッカー + です。休みの日は

↓ 1秒

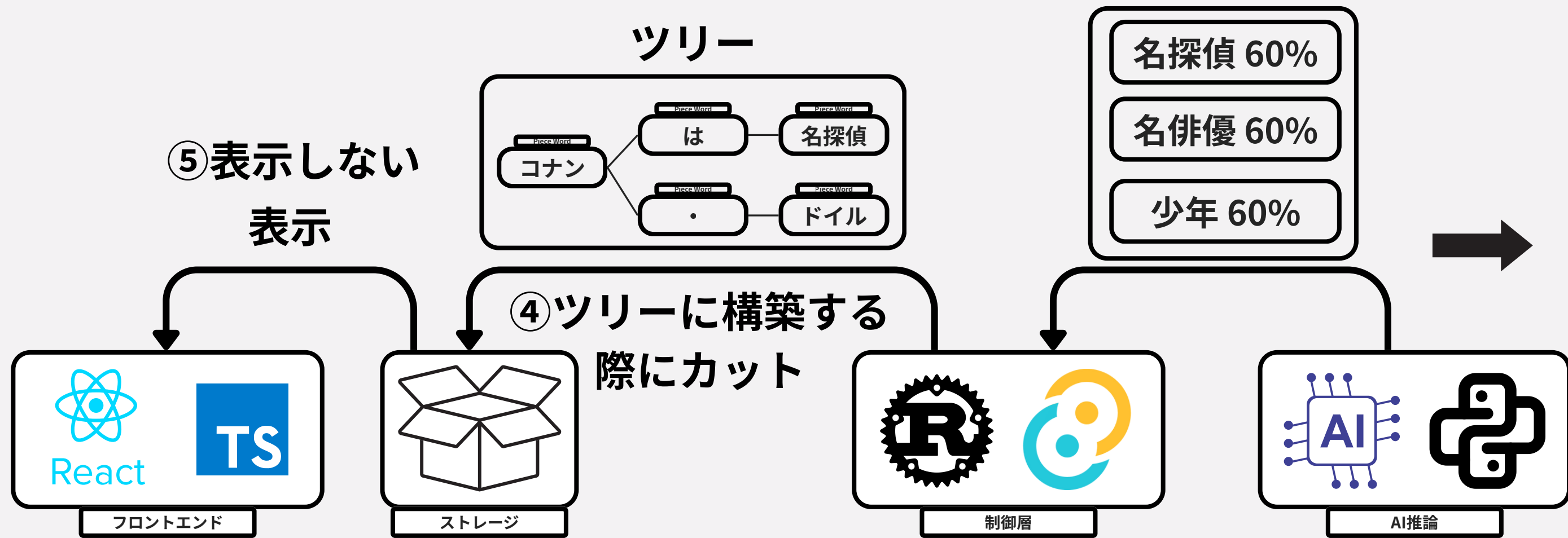
サッカーをします

生成途中のCacheを再利用できる。同じ文章の続きを何度も生成するLLMViewではかなりの高速化につながる。



デメリット
APIを使えない
自分のサーバーが必要

【セーフワード / ブラックリストワード】
データ



不適切なワード
をカット可能

現状は③で実装中
現在の主な用途は

- ・ AIのタグ等を検知→強制終了
- ・ AIのタグを検知→表示しない

APPENDIX

UIについて

The screenshot shows the LLMView AI visualization tool interface. The main area displays a flowchart with four nodes: a root node (80.0%), a node labeled 'ようこそ' (90.0%), a node labeled 'これは' (95.0%), and a final node (90.0%). A search panel on the right is open, showing search results for the character 'る'. The search results include 'ループ', 'ループ後', 'ルート', and 'Tabキーで選択'. A vertical toolbar on the right contains icons for chat, generation (x10), add, search, and display toggle. A yellow callout box on the right lists the following actions: '上から順番に' (In order from top), 'チャット欄を開く' (Open chat bar), '10個生成' (Generate 10 items), 'ノード追加' (Add node), '検索' (Search), and '表示切り替え' (Toggle display).

LLMView AI可視化ツール

新しいプロンプトを入力

新しいセッション作成

セッション: LLMViewへようこそ! このアプリケーションは、AIの言語生成プロセスを可視化するツールです。

LLMViewへようこそ! このアプリケー
2024/1/1 9:00:00

設定

10回生成でツリーを広げやすく

大きなツリーでも検索で好きなノードに簡単に飛べる

ルートから選択ノードまで: LLMView → へ → ようこそ → これは → AIの生成プロセスを可視化するツールです。

構築された文章:
へようこそこれはAIの生成プロセスを可視化するツールです。

検索結果 (3件)

る

ループ
ループ後
ルート カタカナ 選手 → になる
Tabキーで選択

になる
78.8%・サッカー → 選手 → になる

になる

上から順番に

- ・チャット欄を開く
- ・10個生成
- ・ノード追加
- ・検索
- ・表示切り替え