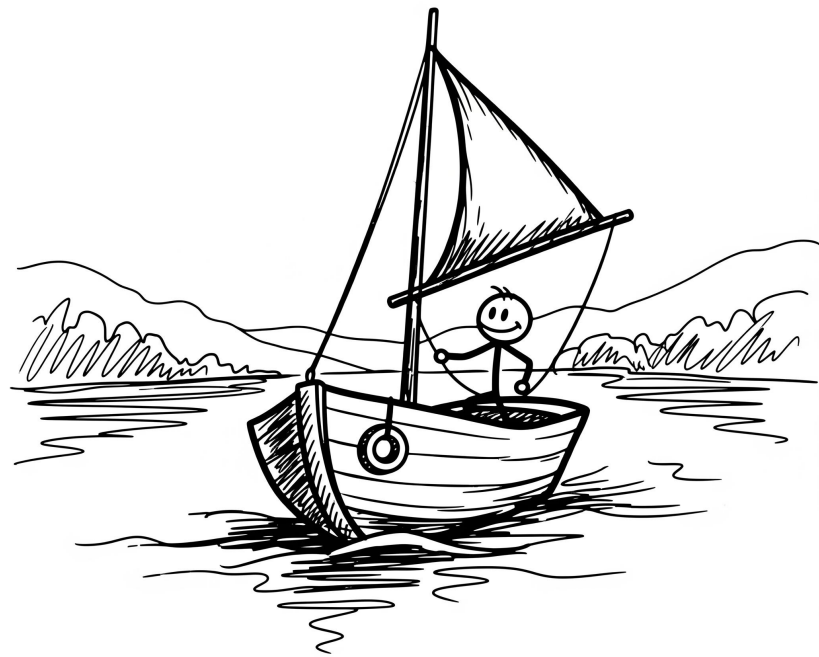


REJS PO SYNTETYCZNYCH DANYCH - OD PERSONY PO TOOLS CALLING



ZNAMY SIĘ MAŁO ...



[HTTPS://WWW.YOUTUBE.COM/@QOoba](https://www.youtube.com/@QOoba)



[HTTPS://GITHUB.COM/QOoba](https://github.com/QOoba)



[HTTPS://WWW.LINKEDIN.COM/IN/QOoba](https://www.linkedin.com/in/QOoba)



TOOLS CALLING

**“Z TYCH NASZYCH ROZMÓW WYŁANIA SIĘ IDEA
WYSTĘPÓW I... JAKICH JESZCZE NIE BYŁO.
STWORZENIA CZEGOŚ ZUPEŁNIE NOWEGO.”**



TOOLS CALLING

```
{"role": "system", "content": "tools_list: create_order(args), search_orders(args) ..." },  
{"role": "user", "content": "Czy możesz znaleźć moje zamówienie, koszulę" }
```

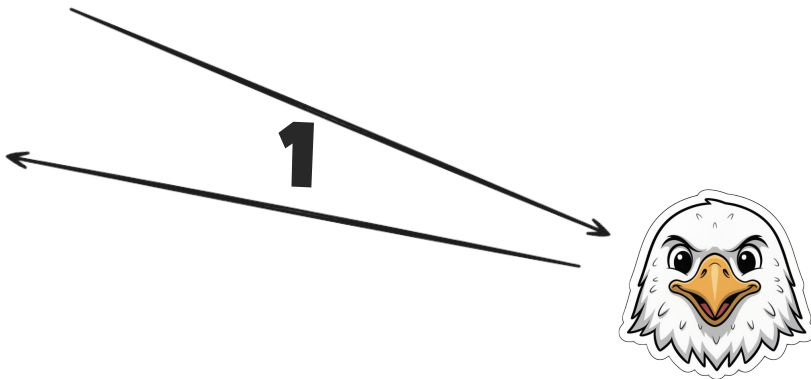
1



TOOLS CALLING

```
{"role": "system", "content": "tools_list: create_order(args), search_orders(args) ..." },  
{"role": "user", "content": "Czy możesz znaleźć moje zamówienie, koszulę" }
```

```
{"role": "assistant", "tool_calls":  
  [{"function": {  
    "name": "search_order",  
    "arguments": {"q": "koszula", ...}  
  }}]  
}
```

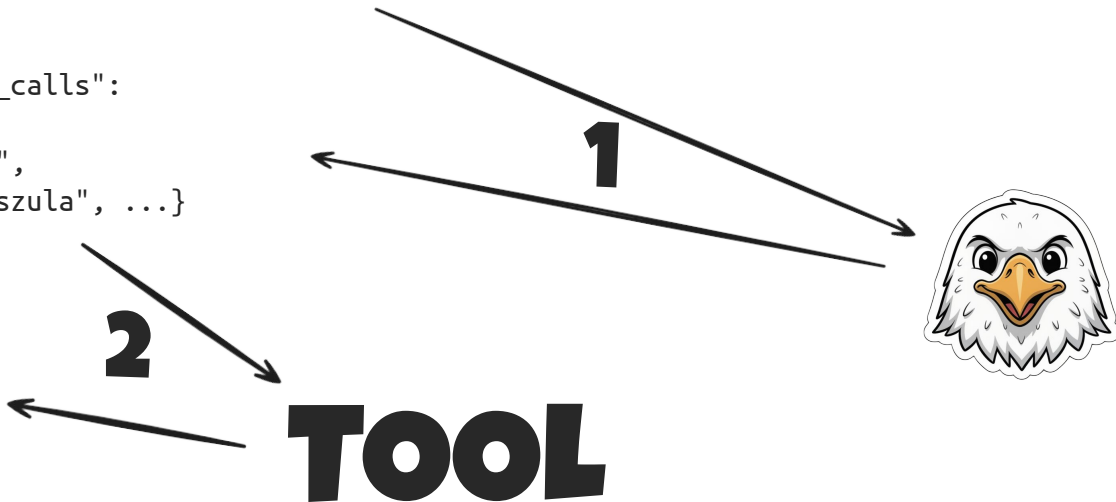


TOOLS CALLING

```
{"role": "system", "content": "tools_list: create_order(args), search_orders(args) ..."} ,  
{"role": "user", "content": "Czy możesz znaleźć moje zamówienie, koszulę" }
```

```
{"role": "assistant", "tool_calls":  
  [{"function": {  
    "name": "search_order",  
    "arguments": {"q": "koszuła", ...}  
  }}]  
}
```

```
{"orders": [{ "id": 1,  
  "category": "koszuła",  
  ... } ] }
```



TOOLS CALLING

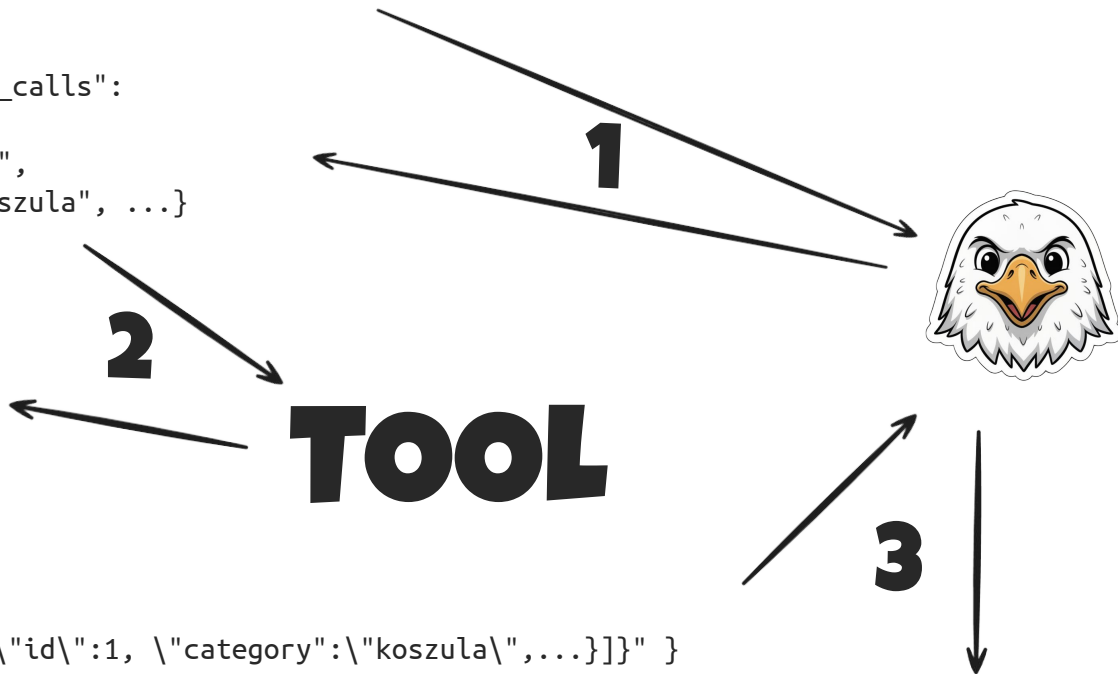
```
{"role": "system", "content": "tools_list: create_order(args), search_orders(args) ..."} ,  
{"role": "user", "content": "Czy możesz znaleźć moje zamówienie, koszulę" }
```

```
{"role": "assistant", "tool_calls":  
  [{"function": {  
    "name": "search_order",  
    "arguments": {"q": "koszuła", ...}  
  }}]  
}
```

```
{"orders": [{ "id": 1,  
  "category": "koszuła",  
  ... } ] }
```

```
{"role": "tool",  
"content": "{ \"orders\": [{ \"id\": 1, \"category\": \"koszuła\", ... } ] }" }
```

```
{"role": "assistant",  
"content": "Tak, znalazłem zamówienie ..." }
```



JAK TO WIDZI BIELIK ?

**“KOŃ... KROWA, KURA, KACZKA... KURA,
KACZKA, DRÓB... (...) O! JEST! WIDZĘ! DROGA...
CHYBA NA OSTROŁĘKĘ.”**





```
[  
  {"role": "user", "content": "Szukam nowego smartfona z 16 gb ramu. \  
    Chciałbym na to przeznaczyć od 2000 do 3000 zł. \  
    I jeszcze nowego laptopa od 3000 do 4000 zł"},  
  
  {"role": "assistant", "tool_calls": [  
    {"function": {"name": "search_products", "arguments": {...}}},  
    {"function": {"name": "search_products", "arguments": {...}}}]  
  }  
]
```

+ TOOLS

```
[
  {"role": "user", "content": "Szukam nowego smartfona z 16 gb ramu. \
    Chciałbym na to przeznaczyć od 2000 do 3000 zł. \
    I jeszcze nowego laptopa od 3000 do 4000 zł"},

  {"role": "assistant", "tool_calls": [
    {"function": {"name": "search_products", "arguments": {...}}},
    {"function": {"name": "search_products", "arguments": {...}}}
  ]
}
```

+ TOOLS

CHAT TEMPLATE

```
<s><|im_start|>system
You are provided with tool signatures that you can use to assist with the user's query. You do not have to use
a tool if you can respond adequately without it. Do not make assumptions about the values to use in tool calls.
If the user's message is missing required parameters or if you do not have an appropriate tool to fulfill the
request, inform the user or ask for clarification instead of attempting to call any tools.

If you decide to invoke a tool, you MUST use the following JSON format:
`<tool_call>{"name": <tool-name>, "arguments": <args-dict>}</tool_call>`

Below is a list of tools in JSON format that you can invoke:

{
  "type": "function",
  "name": "search_products",
  "description": "Wyszukaj produkty w e-bazarze według frazy, kategorii i zakresu cen.",
  "parameters": {...},
  "strict": true
}
<|im_end|>
```

```
[
  {"role": "user", "content": "Szukam nowego smartfona z 16 gb ramu. \
                               Chciałbym na to przeznaczyć od 2000 do 3000 zł. \
                               I jeszcze nowego laptopa od 3000 do 4000 zł"},
  {"role": "assistant", "tool_calls": [
    {"function": {"name": "search_products", "arguments": {...}}},
    {"function": {"name": "search_products", "arguments": {...}}}
  ]
}
```

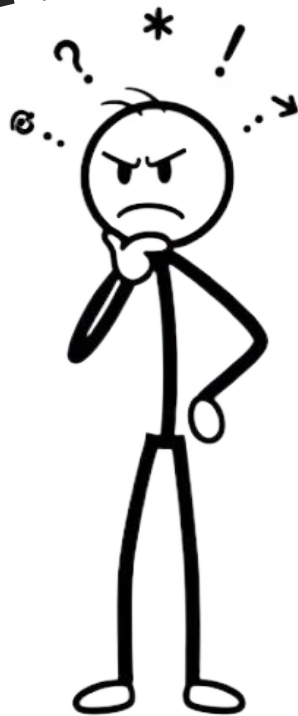
+ TOOLS

CHAT TEMPLATE

```
<|im_start|>user
Szukam nowego smartfona z 16 gb ramu. Chciałbym na to przeznaczyć od 2000 do 3000 zł. I jeszcze nowego laptopa
od 3000 do 4000 zł<|im_end|>
<|im_start|>assistant
<tool_call>{"name": "search_products", "arguments": {"query": "nowy smartfon 16gb ram", "min_price": 2000,
"max_price": 3000}}</tool_call>
<tool_call>{"name": "search_products", "arguments": {"query": "nowy laptop", "min_price": 3000, "max_price":
4000}}</tool_call>
<|im_end|>
```

REALNE FUNKCJE

MA PAN DANE ? A SKĄD MAM MIEĆ ?



REALNE FUNKCJE

Datasets: Locutusque/function-calling-chatml like 172

Modalities: Text Formats: parquet Size: 100K - 1M Libraries: Datasets pandas Croissant +1 License: apache-2.0

Dataset card **Data Studio** **Files and versions** **Community**

Dataset Viewer Auto-converted to Parquet API Embed Data Studio

Split (1)
train · 113k rows

Search this dataset

system_message	function_description	conversations
string · classes	string · lengths	list · lengths
2 values	5.86k	23
You are a helpful assistant with access to the following functions...	<code>{ "name": "get_exchange_rate", "description": "Get the exchange rate..." }</code>	<code>[{ "from": "system", "value": "You are a helpful assistant with access to the followin..." }</code>
You are a helpful assistant with access to the following functions...	<code>{ "name": "get_news_headlines", "description": "Get the latest news..." }</code>	<code>[{ "from": "system", "value": "You are a helpful assistant with access to the followin..." }</code>
You are a helpful assistant with access to the following functions...	<code>{ "name": "generate_password", "description": "Generate a random..." }</code>	<code>[{ "from": "system", "value": "You are a helpful assistant with access to the followin..." }</code>
You are a helpful assistant with access to the following functions...	<code>{ "name": "generate_password", "description": "Generate a random..." }</code>	<code>[{ "from": "system", "value": "You are a helpful assistant with access to the followin..." }</code>
You are a helpful assistant with access to the following functions...	<code>{ "name": "calculate_median", "description": "Calculate the median of ..." }</code>	<code>[{ "from": "system", "value": "You are a helpful assistant with access to the followin..." }</code>
You are a helpful assistant with access to the following functions...	<code>{ "name": "calculate_loan_payment", "description": "Calculate the monthly..." }</code>	<code>[{ "from": "system", "value": "You are a helpful assistant with access to the followin..." }</code>

< Previous 1 2 3 ... 1,130 Next >

World's largest public API Hub

7M+ Developers
84K+ APIs in the Hub
8B+ API calls per month

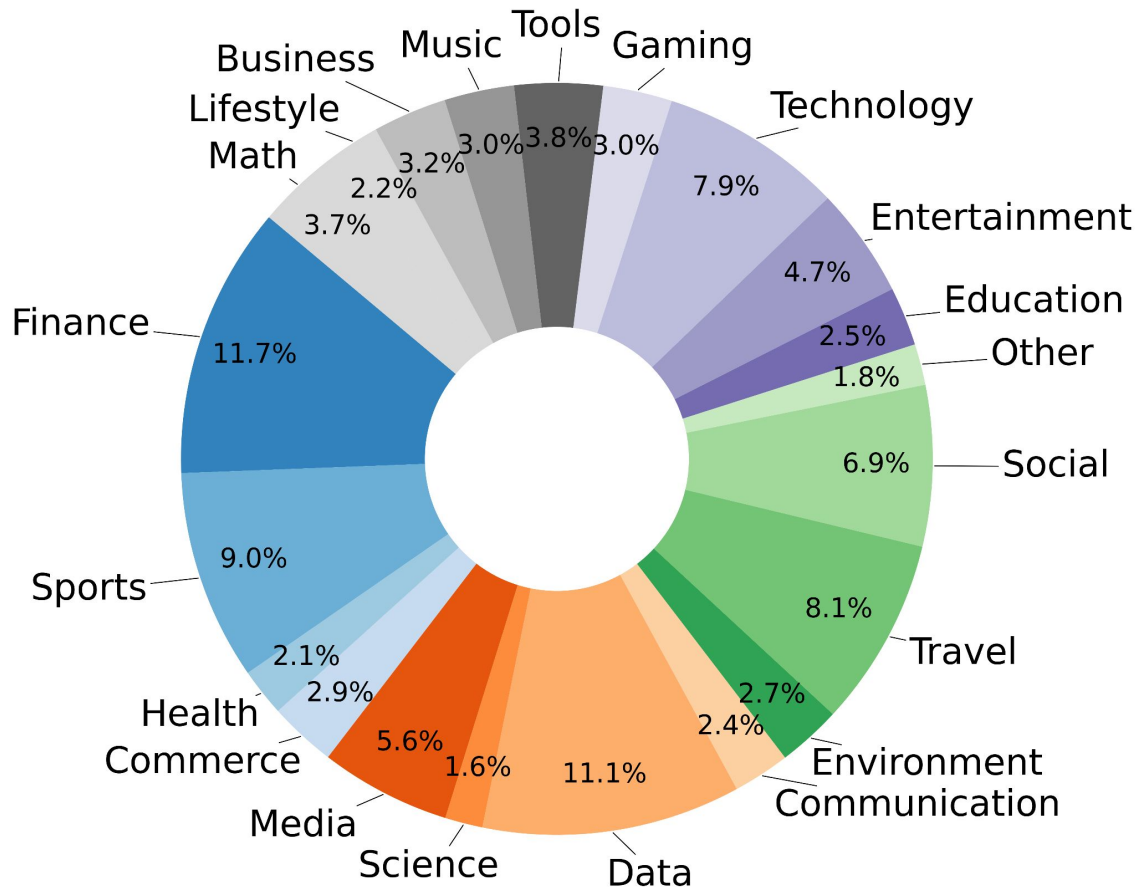
Consume APIs
With indexed search functionality, discovering the perfect API match for your product roadmap is easier than ever.

Manage APIs
Never wonder how many APIs you have or how they're being used. Instead, surface performance and usage patterns instantly and keep a birds-eye view of your API ecosystem.

Publish APIs
Track subscriptions, highlight instructional content. Drive engagement. Monetize APIs

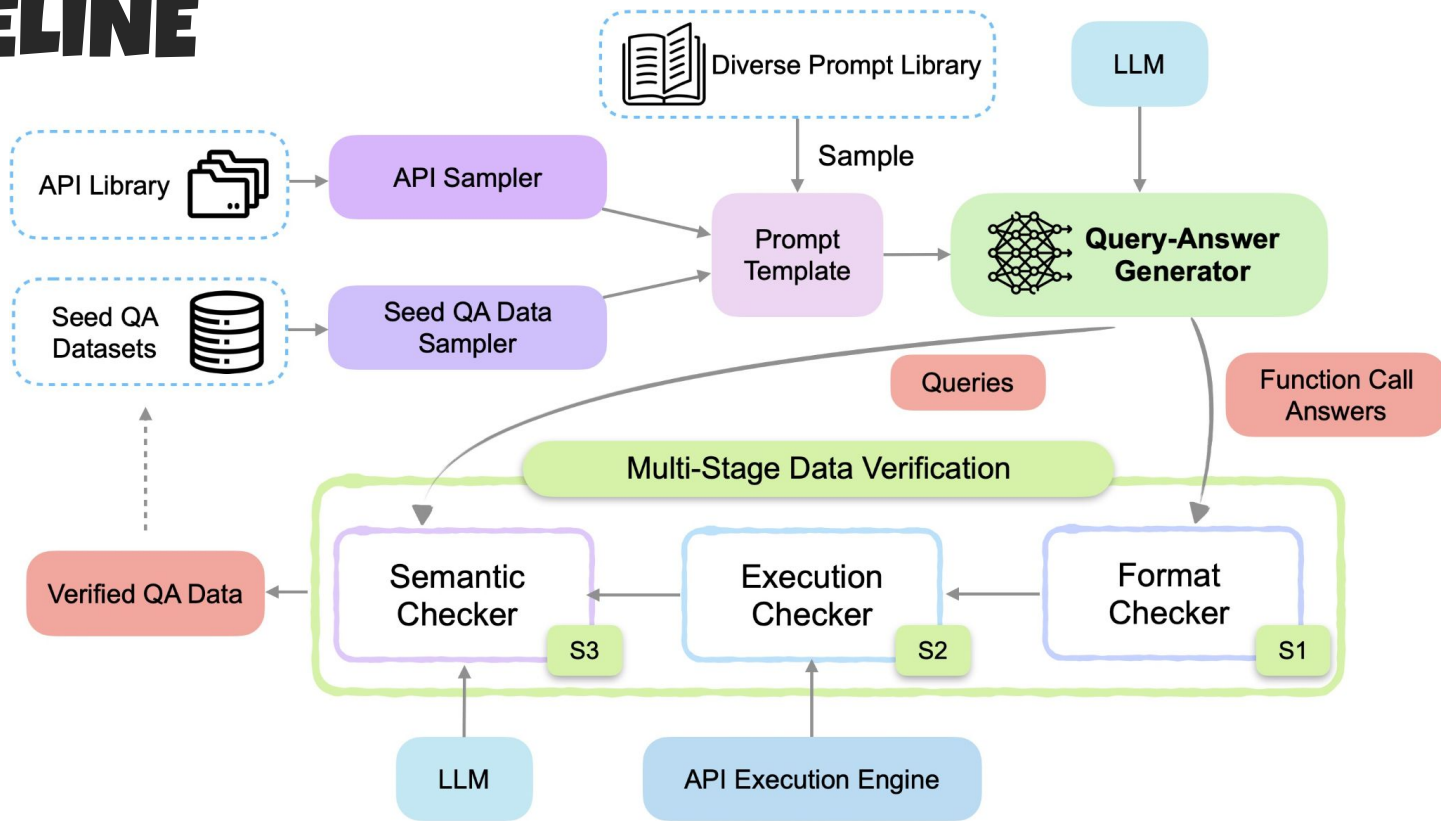
[HTTPS://RAPIDAPI.COM/](https://rapidapi.com/)

[HTTPS://HUGGINGFACE.CO/DATASETS/LOCUTUSQUE/FUNCTION-CALLING-CHATML](https://huggingface.co/datasets/locutusque/function-calling-chatml)



[HTTPS://APIGEN-PIPELINE.GITHUB.IO/](https://apigen-pipeline.github.io/)

PIPELINE



**APIGen: AUTOMATED PIPELINE FOR GENERATING VERIFIABLE AND DIVERSE
FUNCTION-CALLING DATASETS - ARXIV:2406.18518V1**

KROK PO KROKU

TOOL



PYTANIE



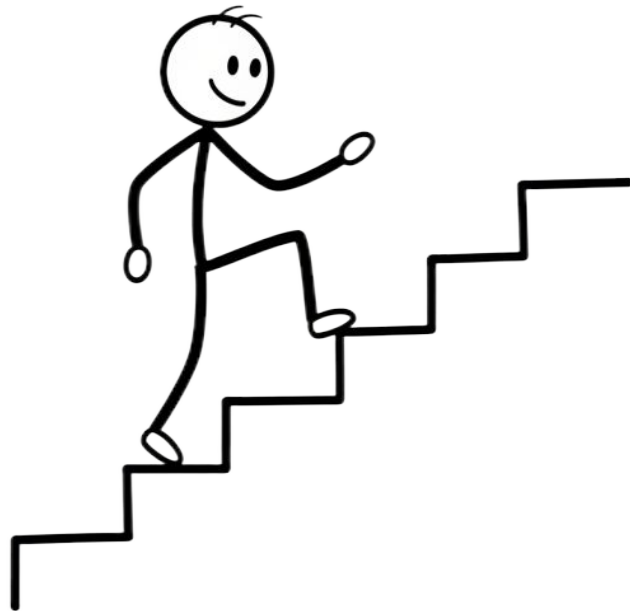
TOOL CALL



**TOOL
RESPONSE**



ODPOWIEDŹ



RÓŻNORODNE DANE

**“NUDA... NIC SIĘ NIE DZIEJE...
DIALOGI NIEDOBRE. W OGÓLE
BRAK AKCJI JEST”**

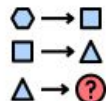


PERSONY

Create {data} with
{persona}



a math problem



a logical reasoning problem



a user prompt to an LLM

1,000,000,000 personas from Persona Hub



a moving company
driver

John, a moving company driver, needs to deliver furniture to three locations. The distances are: 50 miles to the first location; 70 miles to the second location; 80 miles to the third location. John's truck gets 20 miles per gallon and has a 15-gallon tank.

Will he need to refuel during the trip?

You are a moving company driver with a truck that has limited space. You need to load a large couch, a tall bookshelf, and a wide dining table. The couch must be laid flat and loaded first, the bookshelf must stay upright, and the dining table must be the last item loaded.

How do you arrange the items in the truck?

Could you provide step-by-step instructions for safely moving a grand piano from a third-floor apartment to a ground-floor location?

Please include all necessary equipment and best practices for navigating stairs, ensuring both the piano and property remain undamaged during the move.



a chemical kinetics
researcher

Dr. Smith, a chemist, is studying a reaction where compound X decomposes into products Y and Z. The reaction follows first-order kinetics with a rate constant k of 0.5 min^{-1} .

If the initial concentration of compound X is 1.0 M , how long will it take for the concentration of X to decrease to 0.25 M ?

You are analyzing the spatial arrangement of molecules in a reaction chamber. There are three types: A, B, and C. Molecule A is always adjacent to B, but never to C. Molecule B can be adjacent to both A and C.

If molecule C is surrounded by other molecules, which ones must be present around it?

Investigate the effect of temperature on the rate constant of H_2O_2 decomposition with MnO_2 catalyst. Determine the activation energy (E_a) and pre-exponential factor (A) using the Arrhenius equation.

Provide an experimental procedure and data analysis method to extract these parameters.



a musician interested
in audio processing

A musician is studying an audio signal composed of two sine waves. The audio signal $f(t)$ is given by:

$$f(t) = \sin(2\pi \cdot 440t) + \sin(2\pi \cdot 660t)$$

Determine the period of this combined audio signal $f(t)$.

You are setting up a surround sound system with five speakers labeled A, B, C, D, and E. To achieve the best spatial audio effect, you need to place them equidistantly in a circular arrangement.

If speaker A is directly opposite speaker D, who is directly opposite speaker B?

Can you briefly explain the differences between various types of audio filters, such as low-pass, high-pass, band-pass, and notch filters? Additionally, could you please provide examples of how each type of audio filters can be used creatively in music production to enhance or modify sound?

1,000,000,000 synthetic instances created by LLMs

ARXIV 2406.20094

PERSONY PL

SPEAKLEASH DATASET



PERSONY PL

**SPEAKLEASH
DATASET**



LLM

PERSONY PL

**SPEAKLEASH
DATASET**



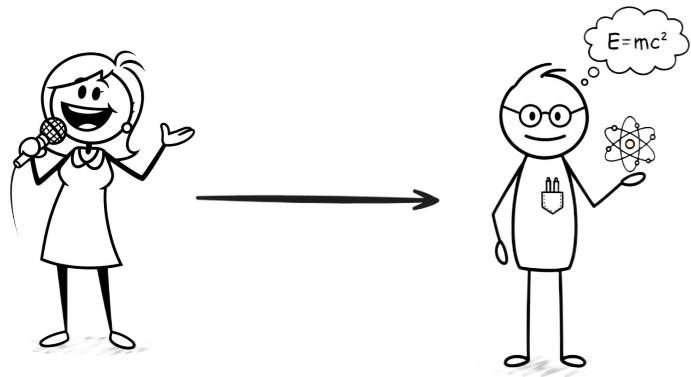
LLM



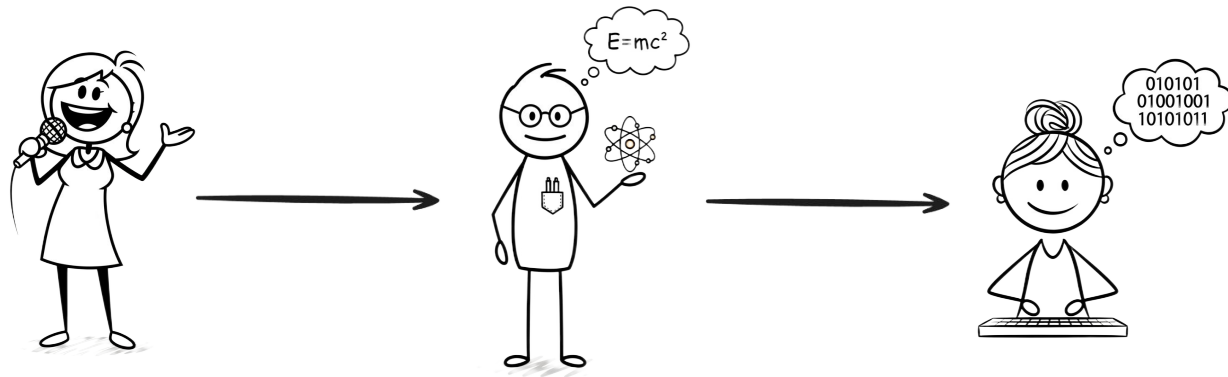
PERSONA 2 PERSONA



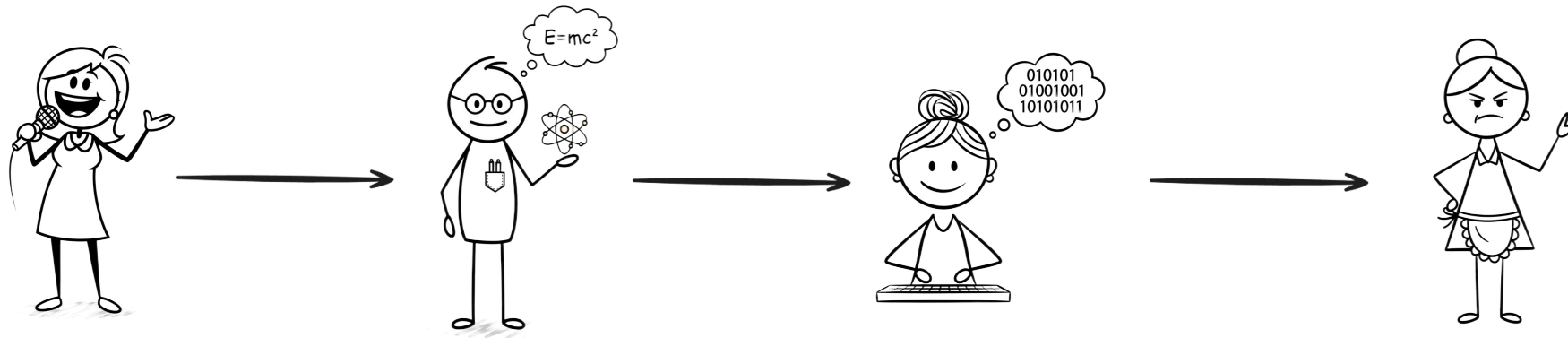
PERSONA 2 PERSONA



PERSONA 2 PERSONA



PERSONA 2 PERSONA



PERSONY PL

{"persona": "Małgorzata - specjalistka ds. dofinansowań w branży filmowej, z pasją śledząca procesy związane z produkcją filmową i scenariuszem."}

{"persona": "Zbigniew, entuzjasta lokalnej społeczności i przyrody, zaangażowany nauczyciel biologii z Szczawnicy"}

{"persona": "Anna, miłośniczka designu i gospodyni domowa ceniąca sobie elegancję i praktyczne rozwiązania w kuchni"}

{"persona": "Klaudia, biegaczka półprofesjonalna dbająca o zrównoważoną dietę i zdrowy styl życia"}

TOOL + PERSONA

search_recipe: Jakie przepisy na zdrowe i smaczne posiłki dla psów mogę znaleźć, uwzględniając składniki takie jak kurczak, marchewka i ryż, preferując lekkostrawną kuchnię bezglutenową?

order_food: Chciałbym zamówić obiad z restauracji "Pod Wawelem" dla mnie i mojego psa. Poproszę pierogi ruskie i zupę pomidorową. Dostawę proszę zrealizować pod adres Planty 12 w Krakowie.

play_music: Czy możesz zagrać mi kołysankę "Ach śpij kochanie" z mojej ulubionej playlisty o nazwie "Kołysanki"?

SCENARIUSZE

**“JAK NIE MA TREŚCI TO PO PROSTU
MOŻNA TAM SOBIE PODKŁADAĆ
RÓŻNE TREŚCI.”**



SINGLE TURN	Question $\Rightarrow$ Tool Call
PARALLEL	Question $\Rightarrow$ Tool Call 1 $\Rightarrow$ Tool Call 2
IRRELEVANT	Question $\Rightarrow$ No Call
CLARIFICATION FLOW	Question $\Rightarrow$ Ask $\Rightarrow$ Info $\Rightarrow$ Tool Call
MULTI STEP	Question 1 $\Rightarrow$ Tool Call $\Rightarrow$ Question 2 $\Rightarrow$ Tool Call

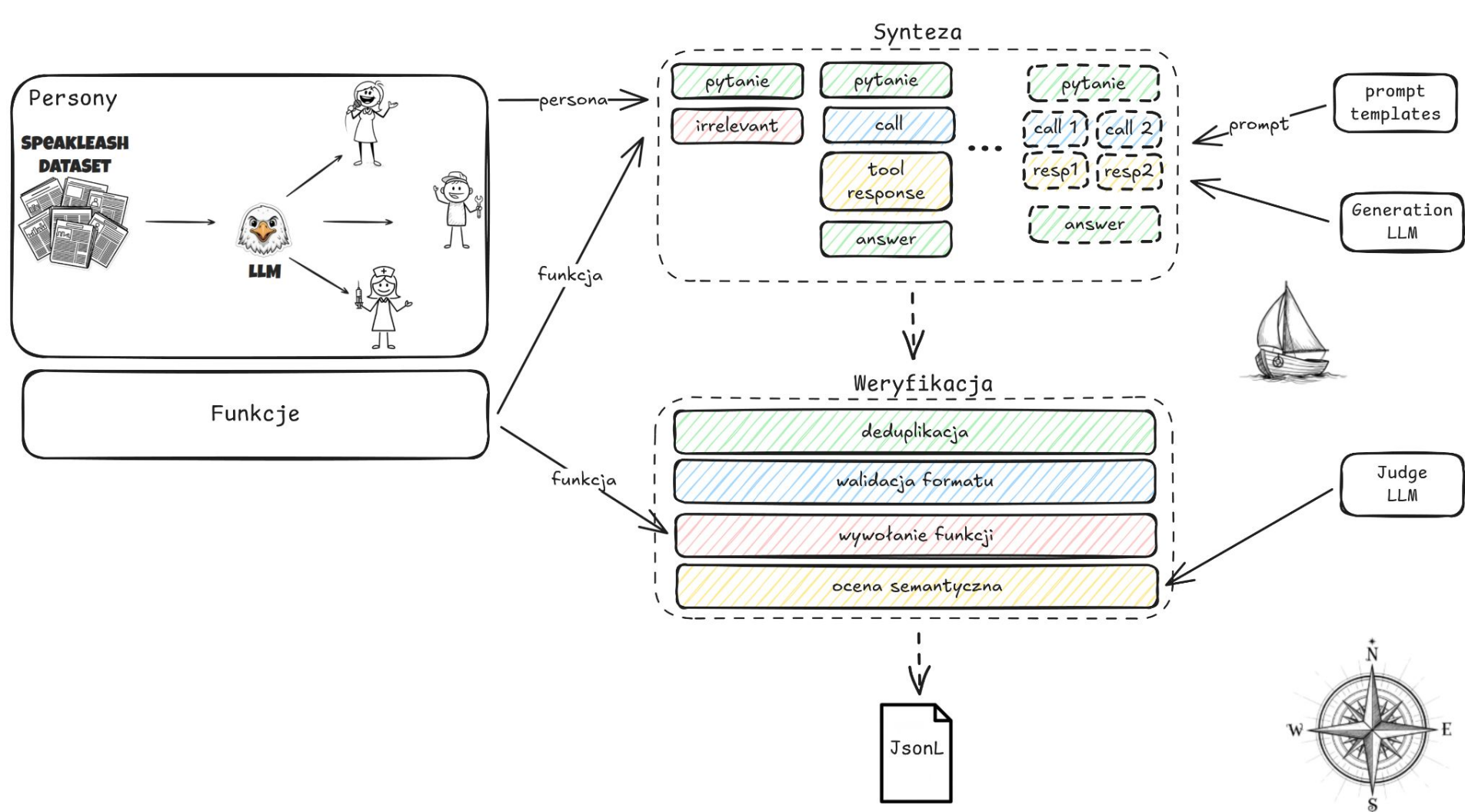
```
(Pipeline("zersoshot")
    .with_workers(5)
    # LLMs
    .with_llm_api("bielik", bielik_url, api_key, bielik_model) # GENERATOR
    # TEMPLATES
    .with_templates("../templates/zeroshot")
    # DATASETS
    .with_dicts_dataset("categories", CATEGORIES)
    # TOOLS DEFINITION
    .with_tools_dataset("tools", [search_orders, get_order_status, create_order, update_order])
.iter_range(1000)
    # SAMPLES
    .sample("personas", 1, "persona")
    .sample_tools("tools", 1, "tools")
    .sample("categories", 1, "categories")
    # QUESTION
    .generate_structured(template="question.j2", llm="bielik", output="q", response_format=QuestionResponse)
    .add_column("question", "q.question")
    # CALL
    .generate_json(template="call.j2", llm="bielik", output="call_arguments")
    .render_tool_call(tool="tool.name", arguments="call_arguments", output="tool_call")
    .render_conversation(conversation="""@user:question|@assistant:tool_calls([tool_call])""", tools="tools",
output="conversation")
    # WRITE
    .write_jsonl("../output/zeroshot.jsonl", value="conversation")
    .run())
```

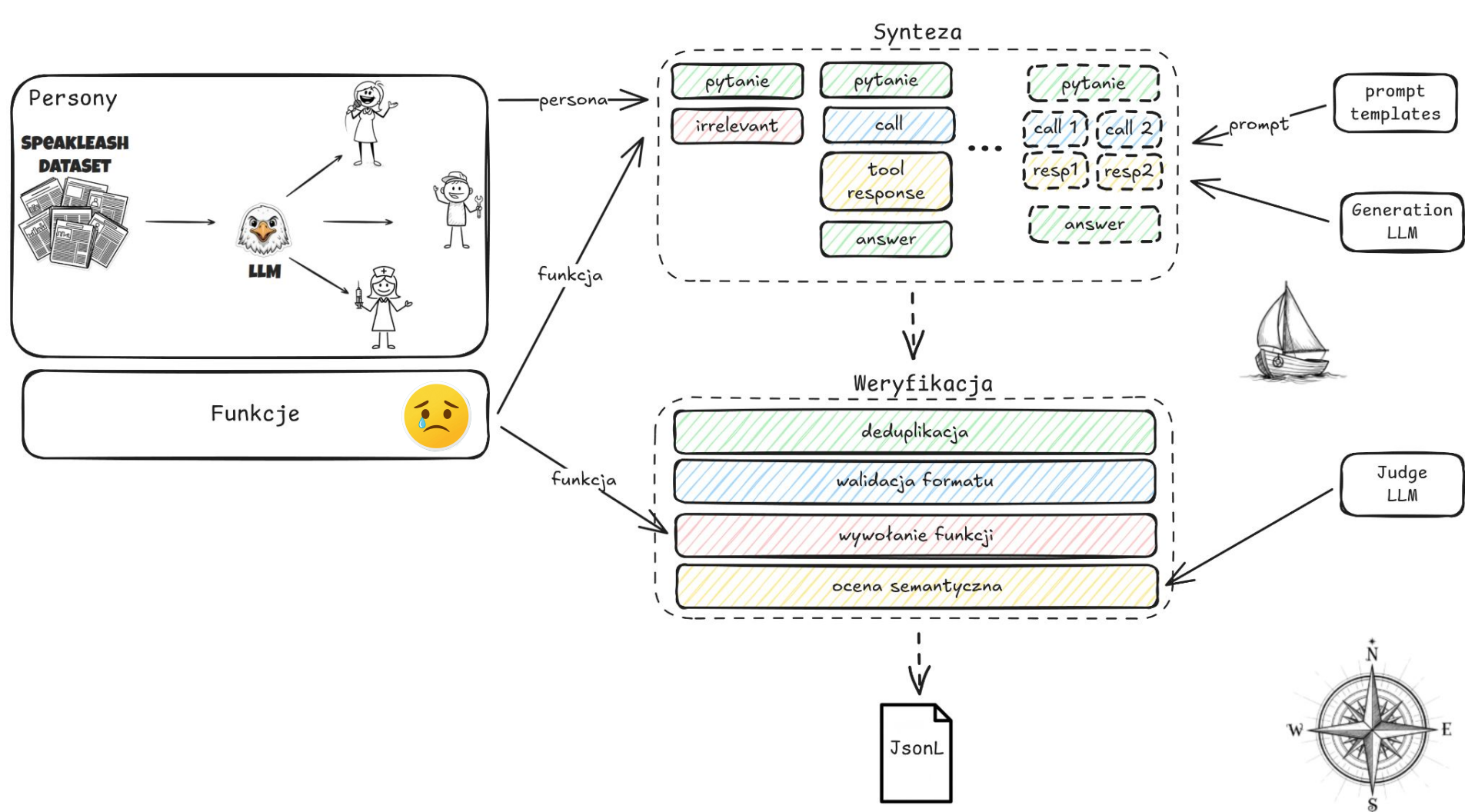
JAKOŚĆ DANYCH

**“EWENTUALNA KRYTYKA MOŻE BYĆ, TAK MUSIMY
ZROBIĆ, ŻEBY TEJ KRYTYKI NIE BYŁO. TYLKO APLAUZ I
ZAAKCEPTOWANIE. TYCH NASZYCH, PRAWDA, PUNKTÓW,
KTÓRE STWORZYMYS”**

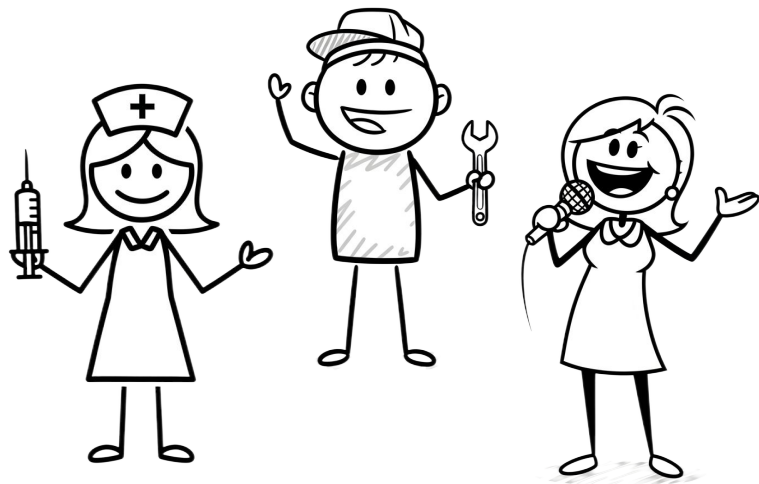


```
(Pipeline("zersoshot")
    .with_workers(5)
    # LLMs
    .with_llm_api("bielik", bielik_url, api_key, bielik_model)      # GENERATOR
    .with_llm_api("deepseek", deepseek_url, api_key, deepseek_model) # JUDGE ✓
    # TEMPLATES
    .with_templates("../templates/zeroshot")
    # DATASETS
    .with_dicts_dataset("categories", CATEGORIES)
    # TOOLS DEFINITION
    .with_tools_dataset("tools", [search_orders, get_order_status, create_order, update_order])
    .iter_range(1000)
    # SAMPLES
    .sample("personas", 1, "persona")
    .sample_tools("tools", 1, "tools")
    .sample("categories", 1, "categories")
    # QUESTION
    .generate_structured(template="question.j2", llm="bielik", output="q", response_format=QuestionResponse)
    .add_column("question", "q.question")
    .check_simhash(input="question", treshold=1) # ✓
    .check_hash(input="question")               # ✓
    # CALL
    .generate_json(template="call.j2", llm="bielik", output="call_arguments")
    .map(evaluate_tool) # ✓
    .validate_json(schema="tool.parameters", instance="call_arguments") # ✓
    .render_tool_call(tool="tool.name", arguments="call_arguments", output="tool_call")
    .render_conversation(conversation="""@user:question|@assistant:tool_calls([tool_calls([tool_call]))""", tools="tools",
output="conversation")
    # JUDGE
    .judge_conversation("conversation", "deepseek", "conversation_judge", "pl") # ✓
    .filter(condition="conversation_judge.intent_alignment > 3 and conversation_judge.tool_choice_accuracy > 3
and conversation_judge.argument_accuracy > 3 and conversation_judge.safety > 3") # ✓
    # WRITE
    .write_jsonl("../output/zeroshot.jsonl", value="conversation")
    .run())
```





PERSONA 2 TOOL



`create_recipe(args)`

`order_car_parts(args)`

`find_song(args)`



```
{'persona': 'Anna, analityk marketingu internetowego, śledząca najnowsze trendy w reklamie online.'}
```



PERSONA 2 TOOL

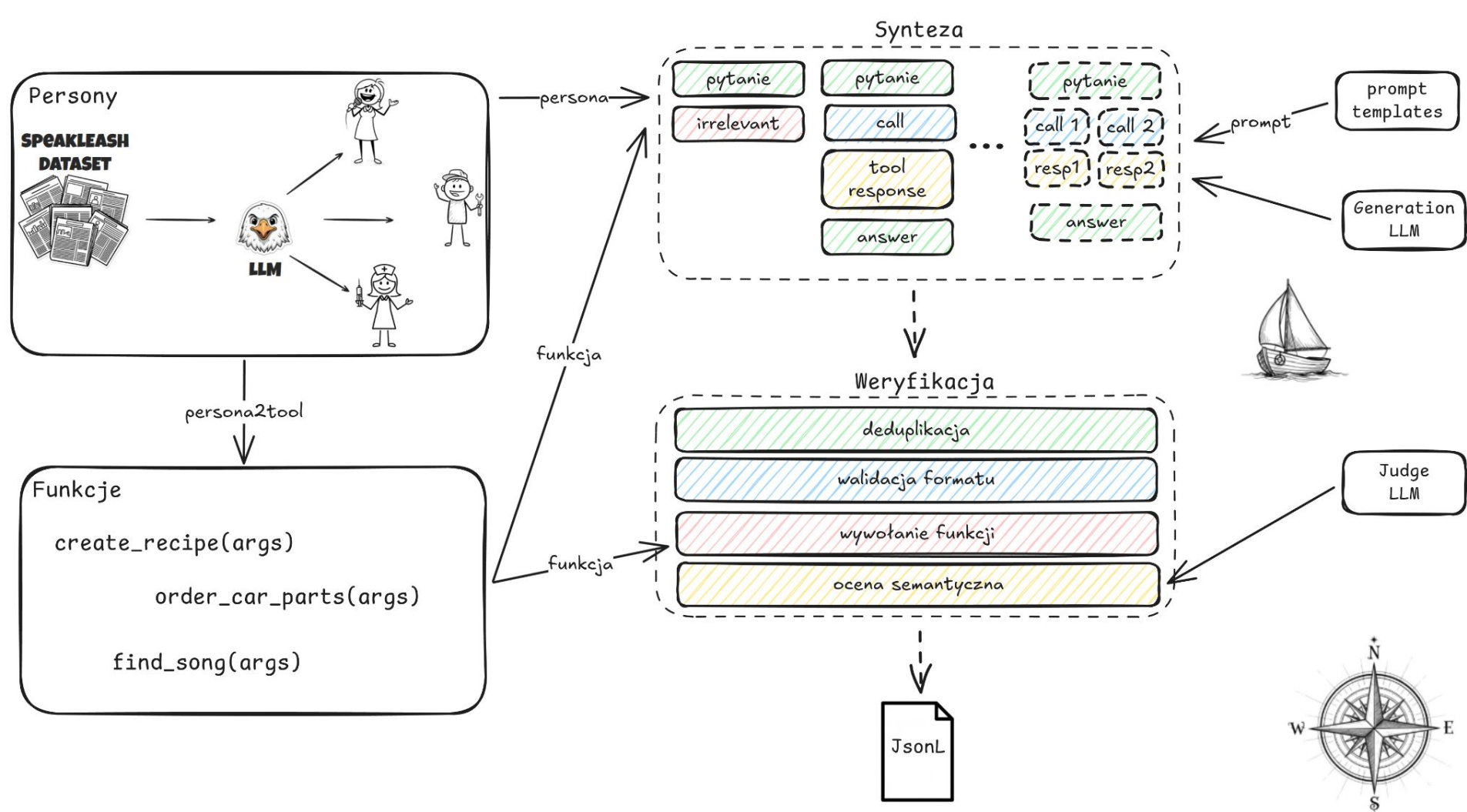


```
{'type': 'function',  
  'name': 'get_top_trends',  
  'description': 'Pobiera informacje o najnowszych trendach w reklamie online dla określonej kategorii i źródła.',  
  'parameters': {'properties': {  
    'category': {  
      'description': 'Kategoria trendów (np. \'social\', \'video\', \'programmatic\')', 'type': 'string'},  
    'source': {  
      'description': 'Źródło danych (np. \'Google\', \'Facebook\', \'Twitter\')', 'type': 'string'},  
    'date_range': {  
      'description': 'Zakres dat (np. \'2023-01-01 to 2023-12-31\')', 'type': 'string'}  
    },  
  'required': ['category', 'source', 'date_range'],  
  'type': 'object',  
  'additionalProperties': False},  
  'strict': True}
```

```
{'persona': 'Kamil - specjalista ds. technologii oczyszczania ścieków z firm, posiadający wieloletnie doświadczenie w projektowaniu i optymalizacji oczyszczalni, ekspert w dziedzinie złoży biologicznych'}
```

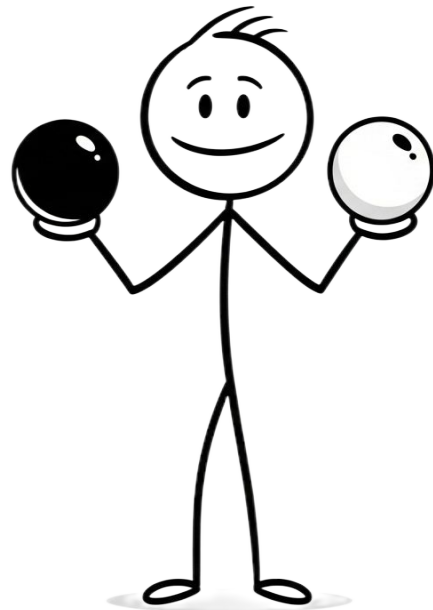
PERSONA 2 TOOL

```
{'type': 'function',  
  'name': 'calculate_bioreactor_efficiency',  
  'description': 'Oblicza efektywność bioreaktora w procesie oczyszczania ścieków.',  
  'parameters': {  
    'properties': {  
      'flow_rate': {  
        'description': 'Przepływ ścieków [m³/d]', 'type': 'number'},  
      'biomass_concentration': {  
        'description': 'Koncentracja biomasy [g/L]', 'type': 'number'},  
      'oxygen_demand': {  
        'description': 'Zapotrzebowanie na tlen [mg O₂/L]', 'type': 'number'},  
      'temperature': {  
        'description': 'Temperatura [°C]', 'type': 'number'}  
    },  
    'required': ['flow_rate', 'biomass_concentration', 'oxygen_demand', 'temperature'],  
    'type': 'object',  
    'additionalProperties': False,  
    'strict': True  
  }  
}
```



SFT, DPO, GRPO

“SĄ TRZY metody GŁOSOWANIA. PIERWSZA PRZEZ APLAUZ ... DRUGA metoda... KULKAMI. JEST TRZECIA metoda PRZEZ PODNIESIENIE RĄK. TA metoda JEST NAJDOSKONALSZA.”



DPO

```
{ "messages": [
  {
    "content": "Hej! Zastanawiam się, czy możesz zaplanować dla mnie i moich 10 przyjaciół super ekologiczny dzień z takimi aktywnościami jak sadzenie drzew, warsztaty zero waste i sprzątanie parku? Mamy na to około 4 godzin i budżet 200 złotych - dasz radę coś fajnego wymyślić?",
    "role": "user"
  }
],
✓ "chosen": {
  "content": "<tool_call>{\"name\": \"pln_ec_dy\", \"arguments\": {\"activities\": [\"sadzenie drzew\", \"warsztaty zero waste\", \"sprzątanie parku\"], \"budget\": 200.0, \"duration_hours\": 4, \"participants\": 10}}</tool_call>",
  "role": "assistant"
},
✗ "rejected": {
  "content": "<tool_call>{\"name\": \"pln_ec_dy\", \"arguments\": {\"activities\": [\"sadzenie drzew, warsztaty zero waste i sprzątanie parku\"], \"budget\": 200.0, \"duration_hours\": 4, \"participants\": 10}}</tool_call>",
  "role": "assistant"
},

"tools": [
  {
    "description": "Planuje ekologiczny dzień z aktywnościami przyrodniczymi i społecznymi.",
    "name": "pln_ec_dy",
    "parameters": {
      "additionalProperties": false,
      "properties": {
        "activities": {"description": "Lista preferowanych aktywności ekologicznych", "items": {}, "type": "array"},
        "budget": {"default": 100.0, "description": "Budżet dostępny na wydarzenie", "type": "number"},
        "duration_hours": {"default": 2, "description": "Czas trwania wydarzenia w godzinach", "type": "integer"},
        "participants": {"default": 10, "description": "Liczba uczestników", "type": "integer"}},
        "required": ["activities"], "type": "object", "strict": true, "type": "function", ... ]
  }
]
```

GRPO

```
{
  "messages": [{ "content": "Halo, projektuję nowy park miejski w strefie umiarkowanej na glebie gliniastej z pełnym nasłonecznieniem i potrzebuję roślin głównie do celów rekreacyjnych. Całkowita powierzchnia to 15 000 m², z czego 8 000 m² to tereny zielone, 500 m² zbiorniki wodne i 1 200 m² powierzchni przepuszczalnej. Czy możesz zasugerować odpowiednie rośliny i obliczyć wskaźniki zieleni dla tego projektu?", "role": "user" }],
  "solution": [

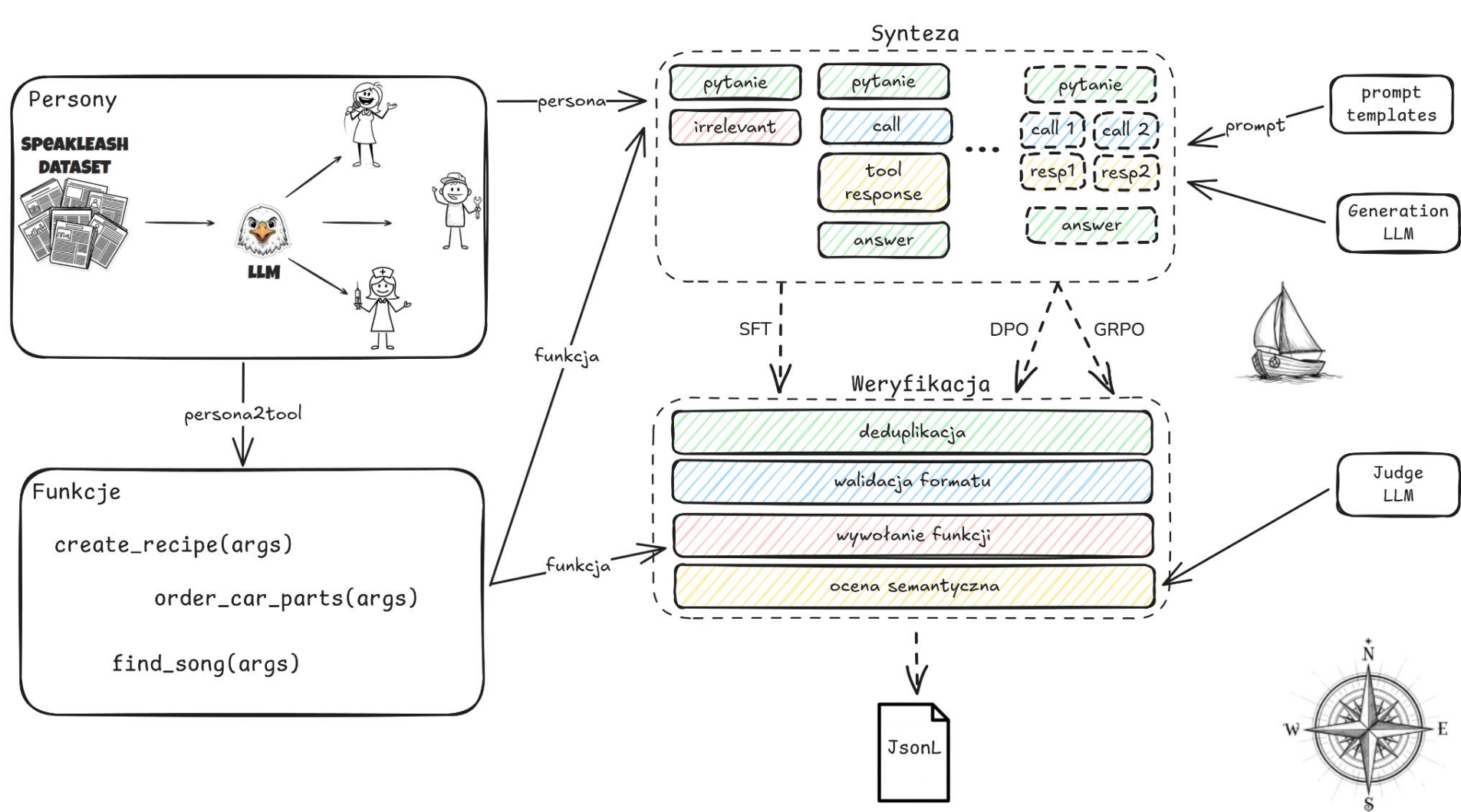
    { "function": { "arguments": {
      "climate_zone": "umiarkowana", "purpose": "rekreacyjne", "soil_type": "gliniasta", "sun_exposure": "pełne" }, "name": "gt_plnt_sgg" } },

    { "function": { "arguments": {
      "area_total": 15000, "green_area": 8000, "permeable_area": 1200, "water_area": 500 }, "name": "clc_grn_rt" } },
  ],

  "validator_id": "tool_use",

  "tools": [
    { "description": "Sugeruje rośliny odpowiednie dla danych warunków i przeznaczenia terenu.", "name": "gt_plnt_sgg", "parameters": { "additionalProperties": false, "properties": { "climate_zone": { "description": "Strefa klimatyczna (np. umiarkowana, śródziemnomorska, kontynentalna)", "type": "string" }, "purpose": { "description": "Przeznaczenie roślinności (ozdobne, rekreacyjne, ekologiczne)", "type": "string" }, "soil_type": { "description": "Typ gleby (np. gliniasta, piaszczysta, próchnicza)", "type": "string" }, "sun_exposure": { "description": "Nasłonecznienie (pełne, częściowe, cień)", "type": "string" } }, "required": [ "climate_zone", "soil_type", "sun_exposure", "purpose" ], "type": "object" }, "strict": true, "type": "function" },

    { "description": "Oblicza wskaźniki zieleni i przepuszczalności dla projektu urbanistycznego.", "name": "clc_grn_rt", "parameters": { "additionalProperties": false, "properties": { "area_total": { "description": "Całkowita powierzchnia terenu w m²", "type": "number" }, "green_area": { "description": "Powierzchnia terenów zielonych w m²", "type": "number" }, "permeable_area": { "default": 0.0, "description": "Powierzchnia przepuszczalna w m²", "type": "number" }, "water_area": { "default": 0.0, "description": "Powierzchnia zbiorników wodnych w m²", "type": "number" } }, "required": [ "area_total", "green_area" ], "type": "object" }, "strict": true, "type": "function" },
  ]
}
```





KONIEC ?