**Porównanie środowisk do tworzenia Small Agents**

**1. Llama.cpp**

**Typ:** Środowisko lokalne  
**Boty:** Modele językowe LLaMA, możliwe inne modele językowe, w tym mniejsze modele oparte na transformatorach, mogą być używane do ogólnych zastosowań (od prostych chatbotów po zaawansowane agentów AI).  
**Opis:** Llama.cpp służy do uruchamiania różnych wersji modeli LLaMA oraz innych podobnych modeli. Umiejętność programowania jest wymagana, ponieważ wymaga to konfiguracji i zarządzania kodem. Główną zaletą jest wydajność – można uruchomić modele na słabszych komputerach, nie wymagających GPU. Świetne do lokalnych testów i eksperymentów z różnymi modelami.

**2. Ollama**

**Typ:** Środowisko lokalne   
**Boty:** Modele LLaMA, GPT, BLOOM, Mistral, i inne modele open-source. Możliwość uruchamiania różnorodnych modeli w jednym narzędziu.  
**Opis:** Ollama to środowisko, które umożliwia łatwe uruchamianie wielu modeli AI bez konieczności pisania kodu. Umożliwia łatwą integrację różnych modeli w jednym interfejsie. Głównie nastawione na użytkowników, którzy potrzebują łatwości w użyciu, a nie na optymalizację wydajności jak w przypadku Llama.cpp.

**3. Vertex AI**

**Typ:** Środowisko chmurowe (Google Cloud)  
**Boty:** Modele Gemini (Gemini-1, Gemini-1.5-flash), BERT, PaLM, i inne modele dla przetwarzania języka naturalnego, rozpoznawania obrazów, przetwarzania mowy i analizy tekstu.  
**Opis:** Vertex AI to platforma chmurowa od Google, która umożliwia łatwe tworzenie i uruchamianie zaawansowanych modeli AI. Nie wymaga zaawansowanej wiedzy programistycznej, ponieważ użytkownicy mogą używać gotowych modeli i dostosowywać je za pomocą interfejsów graficznych. Skierowane na szeroką gamę zastosowań AI. Modele mogą być używane do różnych celów, od chatbotów po rozpoznawanie obrazów. Platforma jest skalowalna, więc może obsługiwać duże modele językowe bez potrzeby posiadania mocnego sprzętu lokalnego.

**4. Dialogflow**

**Typ:** Środowisko chmurowe (Google Cloud)  
**Boty:** Chatboty i voiceboty, w tym modele rozpoznawania intencji, słów kluczowych, rozpoznawanie mowy, generowanie odpowiedzi w oparciu o wstępnie zdefiniowane modele.  
**Opis:** Dialogflow to narzędzie Google Cloud stworzone z myślą o łatwym tworzeniu chatbotów i voicebotów. Jest to platforma, która wymaga minimalnych umiejętności programowania, ponieważ obsługuje proces tworzenia agenta AI za pomocą interfejsów graficznych. Jest łatwa w użyciu, ale nie oferuje tak dużych możliwości personalizacji jak bardziej zaawansowane platformy, takie jak Llama.cpp, Ollama, czy Vertex AI. Idealna do prostych chatbotów na stronach internetowych lub do integracji z aplikacjami mobilnymi.

**5. Azure AI**

**Typ:** Środowisko chmurowe (Microsoft)  
**Boty:** Modele OpenAI (GPT-4, GPT-3.5), Custom GPT, i inne modele dla przetwarzania języka naturalnego, obrazów, mowy.  
**Opis:** Azure AI to platforma chmurowa od Microsoftu, oferująca szeroki zestaw narzędzi do sztucznej inteligencji. Można korzystać z gotowych usług, jak również trenować własne modele w Azure Machine Learning. Użytkownicy mogą tworzyć i wdrażać aplikacje oparte na AI z pomocą interfejsów graficznych, co nie wymaga dużych umiejętności programistycznych. Model GPT-4 i GPT-3.5 dostępne w ramach usług, pozwalają na łatwe tworzenie chatbotów i agentów AI. Dodatkowo, Azure AI ma integrację z aplikacjami Microsoftu, co może być przydatne w korporacyjnych rozwiązaniach.

**6. RASA**

**Typ:** Środowisko lokalne (open-source)  
**Opis:** Platforma open-source, co daje pełną kontrolę nad procesem tworzenia agentów AI. Dzięki temu użytkownicy mogą dostosować modele do specyficznych potrzeb, w tym trenować je na własnych danych. Rasa obsługuje wiele integracji z popularnymi platformami, takimi jak Slack czy Facebook Messenger, co pozwala na łatwą implementację chatbotów w różnych środowiskach. Dodatkowo, system jest wielojęzyczny, co umożliwia tworzenie agentów działających w wielu językach. Pomimo swoich zalet, Rasa wymaga od użytkowników pewnej wiedzy technicznej i poświęcenia czasu na konfigurację, co sprawia, że jest mniej przystępna dla początkujących. Dla osób, które oczekują prostych, gotowych rozwiązań, może być zbyt skomplikowana. Ponadto, proces treningu modeli może być czasochłonny i wymagać znacznych zasobów obliczeniowych, co nie każdemu użytkownikowi odpowiada.

**7. Botpress**

**Typ:** Środowisko chmurowe (open-source)  
**Opis:** platforma, która umożliwia tworzenie zaawansowanych chatbotów i agentów AI. Oferuje intuicyjny interfejs, który umożliwia łatwe tworzenie botów bez konieczności programowania, co czyni go przystępnym dla osób bez doświadczenia technicznego. Botpress wspiera integrację z popularnymi kanałami komunikacyjnymi, takimi jak Messenger, WhatsApp czy Slack, co pozwala na szerokie zastosowanie chatbotów w różnych środowiskach. Dodatkowo, platforma oferuje zaawansowane funkcje, takie jak NLP (Natural Language Processing), co umożliwia bardziej złożoną interakcję z użytkownikami. Jednak, mimo łatwości obsługi, bardziej zaawansowane funkcje, takie jak integracje z bazami danych czy API, wymagają wiedzy technicznej. Botpress może także wymagać większych zasobów serwerowych przy intensywnym użytkowaniu. Dla użytkowników poszukujących prostych rozwiązań lub bez wiedzy programistycznej, platforma może okazać się zbyt zaawansowana.

**8. LangChain**

**Typ:** Środowisko lokalne open-source  
**Opis:** Framework (biblioteka open-source) do tworzenia aplikacji wykorzystujących modele językowe (LLM)   
LangChain to zaawansowany framework umożliwiający łatwe rozszerzanie danych w czasie rzeczywistym poprzez integrację modeli językowych (LLM) z różnorodnymi źródłami danych oraz systemami wewnętrznymi i zewnętrznymi. Dzięki rozbudowanej bibliotece integracji LangChain pozwala łączyć modele z dostawcami modeli AI, narzędziami, magazynami wektorowymi oraz wyszukiwarkami, zapewniając płynną interoperacyjność.

Framework umożliwia swobodne zamienianie modeli w trakcie eksperymentów inżynierskich, co ułatwia znalezienie najlepszego rozwiązania dopasowanego do potrzeb aplikacji. Wraz z rozwojem branży i pojawianiem się nowych technologii, LangChain pozwala szybko się dostosowywać — jego elastyczne abstrakcje umożliwiają dynamiczne zmiany bez utraty tempa rozwoju.

Choć LangChain może działać samodzielnie, najlepiej wykorzystać jego potencjał w połączeniu z innymi produktami ekosystemu:

* **LangSmith** – narzędzie do debugowania, monitorowania i oceny agentów. Umożliwia analizę słabo działających przebiegów aplikacji LLM, ocenę trajektorii agentów oraz zwiększanie wydajności w środowisku produkcyjnym.
* **LangGraph** – niskopoziomowy framework do tworzenia agentów, pozwalający budować systemy zdolne do realizacji złożonych zadań. Oferuje wsparcie dla pamięci długoterminowej, przepływów pracy z udziałem człowieka oraz elastyczną, konfigurowalną architekturę.
* **LangGraph Platform** – platforma do wdrażania i skalowania agentów. Ułatwia zarządzanie długotrwałymi, stanowymi przepływami pracy, a także pozwala na szybkie prototypowanie wizualne, współdzielenie agentów w zespołach i przyspieszanie iteracji rozwoju.

**9. LangFlow**

**Typ:** Środowisko lokalne i chmurowe open-source  
**Opis:** Framework (biblioteka open-source) do tworzenia aplikacji wykorzystujących modele językowe (LLM)   
LangFlow to narzędzie do tworzenia i wdrażania agentów oraz przepływów pracy zasilanych sztuczną inteligencją. LangFlow oferuje deweloperom zarówno intuicyjny interfejs wizualny, jak i wbudowany serwer API, który umożliwia przekształcenie każdego agenta w punkt końcowy API, gotowy do integracji z aplikacjami zbudowanymi na dowolnym frameworku czy stosie technologicznym. Narzędzie to wyposażone jest we wszystkie niezbędne funkcje, wspiera główne modele językowe (LLM), bazy danych wektorowe oraz stale rosnącą bibliotekę narzędzi AI.

**Kluczowe funkcje LangFlow:**

* **Wizualny kreator** – Umożliwia szybkie rozpoczęcie pracy i iterowanie nad przepływami AI dzięki łatwemu w użyciu interfejsowi wizualnemu.
* **Dostęp do kodu** – Programiści mogą modyfikować dowolny komponent, dostosowując aplikacje za pomocą Pythona, zapewnia to elastyczność i pełną kontrolę nad projektem.
* **Playground** – Środowisko testowe, które pozwala na natychmiastowe testowanie i iterowanie przepływów z pełną kontrolą nad każdym etapem procesu.
* **Wieloagentowy** – LangFlow umożliwia zarządzanie wieloma agentami.
* **Wdrażanie jako API lub eksport do JSON** – Agenci mogą być łatwo wdrażani jako punkty końcowe API lub eksportowani w formie JSON, co umożliwia ich integrację z aplikacjami opartymi na Pythonie.
* **Obserwowalność i integracja** – LangFlow wspiera monitorowanie i analizowanie przepływów pracy przy użyciu narzędzi takich jak LangSmith i LangFuse.
* **Bezpieczeństwo i skalowalność** – gotowe na potrzeby przedsiębiorstwa.