الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٠08 ماي 1945٠ بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Master**
Filière : **Computer Science**
Specialty : **Système d'Information et Web (SIW)**

Theme

Agentic Paradigms for Reliable LLM Systems

Presented by :

Mansour Houssam

Submission Date: **September, 2025**

In front of the jury composed of:

Pr AMAR BEN SABER Djamel	President
Dr AMRANE Abdelkader	Examiner
Dr Mohammed BEDJAOUI	Supervisor

Academic Year : 2024/2025

Abstract

Language-model–driven agents are increasingly deployed in settings that demand reliability over long horizons. Yet the ecosystem offers many mechanisms with overlapping promises. This thesis synthesizes four dominant paradigms—**tool use**, **planning**, **feedback/learning**, and **memory**—and asks a practical question: *when does each paradigm actually help?* We ground our analysis in a common set of failure modes (hallucination, long-horizon credit assignment, exposure bias, tool/API hallucination, forgetting, and compounding error) and map each paradigm to the conditions under which it mitigates these failures. We synthesize that tool use is most effective for knowledge- and computation-centric tasks (grounded retrieval, code/solver execution, API orchestration); planning helps when problems decompose into verifiable subgoals or require look-ahead and search; feedback/learning raises first-pass quality and enables iterative self-correction or preference alignment; and memory confers continuity, personalization, and cost/latency control across sessions. At the same time, we highlight composition-level gaps: the literature rarely evaluates stacked combinations (e.g., ReAct + RAG + Reflexion + tiered memory) under drift, cost, and safety constraints. We conclude with actionable guidance for designing minimal viable compositions, emphasizing verifiable checkpoints, interface observability, and deliberate latency budgeting. The result is a practitioner-oriented framework that links task properties and failure modes to concrete agentic design choices.

Résumé

Les agents basés sur des modèles de langage sont de plus en plus utilisés dans des contextes exigeant une fiabilité à long terme. Cependant, l'écosystème propose de nombreux mécanismes aux promesses souvent chevauchantes. Ce mémoire synthétise quatre paradigmes dominants — **l'utilisation d'outils**, **la planification**, **le retour d'expérience/apprentissage** et **la mémoire** — et pose une question pratique : *dans quels cas chaque paradigme est-il réellement utile ?* Nous ancrons notre analyse dans un ensemble commun de modes d'échec (hallucination, attribution de crédit à long terme, biais d'exposition, hallucination d'API/outils, oubli et erreur cumulative) et établissons la correspondance entre chaque paradigme et les conditions dans lesquelles il atténue ces défaillances. Nous concluons que l'usage d'outils est le plus efficace pour les tâches centrées sur la connaissance et le calcul ; la planification est bénéfique lorsque les problèmes peuvent être décomposés en sous-objectifs vérifiables ou nécessitent de l'anticipation et de la recherche ; le retour d'expérience et l'apprentissage améliorent la qualité initiale et permettent l'auto-correction ou l'alignement sur les préférences ; enfin, la mémoire assure la continuité, la personnalisation et la maîtrise des coûts et de la latence à long terme. Ce travail propose ainsi un cadre pratique reliant les propriétés des tâches et les modes d'échec aux choix de conception des agents.

الملخص

تُستخدم الوكلاء المعتمدة على النماذج اللغوية بشكل متزايد في البيئات التي تتطلب موثوقية على المدى الطويل. ومع ذلك، فإن النظام البيئي يقدم آليات متعددة ذات وعود متداخلة. يجمع هذا البحث بين أربعة نماذج رئيسية --- **استخدام الأدوات**، **التخطيط**، **التغذية الراجعة/التعلم**، و**الذاكرة** --- ويطرح سؤالًا عمليًا: متى يكون كل نموذج مفيدًا فعلاً؟ نعتمد في تحليلنا على مجموعة من أنماط الفشل الشائعة (كالهلوسة، وإسناد الفضل طويل الأمد، والانحياز في التعرض، وهلوسة الأدوات أو الواجهات البرمجية، والنسيان، والخطأ التراكمي)، ونربط كل نموذج بالظروف التي يخفف فيها هذه الإخفاقات. نستنتج أن استخدام الأدوات هو الأكثر فعالية في المهام القائمة على المعرفة والحساب، وأن التخطيط يساعد عندما يمكن تقسيم المشكلة إلى أهداف فرعية قابلة للتحقق أو تتطلب استشرافًا وبحثًا، وأن التغذية الراجعة والتعلم يحسّنان الجودة الأولية ويمكّنان من التصحيح الذاتي، وأن الذاكرة توفر الاستمرارية والتخصيص والتحكم في التكلفة والزمن عبر الجلسات. يقدّم هذا العمل إطارًا عمليًا يربط بين خصائص المهام وأنماط الفشل وخيارات التصميم في بناء الوكلاء الذكيين.