Mercredi 15 janvier 2025  
  
Récapitulatif des évènements avant les vacances de Noël :

Thierry, Yann, Jérôme et moi travaillons sur le PoC héracles. Après de nombreux échanges avec différents prestataires, nous avons fini par choisir Orange Business comme solution tiers. Ce choix à par ailleurs amené à des modifications quant au projet en tant que tel. Entre les négociations budgétaires, du cahier des charges et du programme, nous en avons conclu que le PoC Héracles n’était peut être pas la priorité n°1 puisque ce cas d’usage a été qualifié de « tiré du chapeau » signifiant que nous nous y sommes penchés dessus uniquement parce que cela avait été mentionné au COMEX comme étant une bonne idée, sans analyses de ROI, ou de cas d’usages parallèles pouvant être plus simple, moins couteux et avec de meilleurs rendus (QUICK WIN).

Cependant, le choix dirigé vers Orange nous a permis plusieurs choses :

* Premièrement, négocier un budget entre les différents partenariats entre Orange Microsoft et Gecina, nous permettant de pouvoir couvrir plusieurs sujets en simultané et ce pour globalement moins cher que le prix du PoC héraclès (et de ses variations/ fournisseurs).
* Second, couvrir l’acculturation et l’apprentissage de copilot au près des collaborateurs : effectivement l’avènement de l’IA dans nos vies pousse la plupart des collaborateurs à utiliser ChatGPT (freemium) dans leurs tâches quotidiennes pour améliorer leurs processus, ce qui semble normal à première vue. Néanmoins, ChatGPT peut utiliser les données utilisateurs pour l’entraînement de ses modèles, exposants ainsi des données Gecina à un usage commerciale ainsi qu’à des failles de sécurité en dehors du ressort de Gecina. Il est donc nécessaire de s’équiper en licences copilot et d’acculturer les collaborateurs à l’outils et l’IA en général.
* De plus, alors que Microsoft était en pleine campagne marketing de leurs avancées en IA avec l’arrivée « d’agents » au sein de leurs produit, une demande provenant du directeur général nous a été livrée, et il nous a été demandé de solutionner la demande en utilisant les produits copilot. Néanmoins une problématique s’impose : pour moi, Microsoft procède à une déformation de la réalité pour ne pas dire du mensonge marketing. Le choix concernant la nomenclature de leurs produits est minutieusement pensé pour surfer sur les trends agentic 2024/2025. Or un agent peut être qualifié d’agent uniquement s’il est composé des 3 briques suivantes : la mémoire, l’appel d’outils/fonctions ainsi qu’une capacité de raisonnement (chain-of-thoughts). Ce qui n’est absolument pas le cas pour les « agents » microsoft copilot, avec lesquels il n’est pas possible de diriger de recherches en utilisant les métadonnées ou une chronologie précise. Ces outils sont centralisés autour de la connectivité de l’EntraID, créant ainsi un effet boîte noire. Ainsi, lorsque la demande nous a été transmise, nous avons essayé de répondre à la problématique en créant un « agent » copilot dont la source d’information était un sharepoint sur lequel nous enregistrons les pièces jointes reçues par le directeur, dans le but de pouvoir les synthétiser et dialoguer avec elles. Néanmoins, l’agent n’ayant pas la capacité de temporaliser ou sélectionner les pièces jointes, il n’était pas possible d’obtenir de synthèse exploitable.
* Ayant initialement automatisé la réception et l’enregistrement de pièce jointe dans un sharepoint, il nous a semblé être une bonne idée d’inspecter les capacités de power automate à répondre à la demande dans son ensemble. Ainsi, dans l’attente de l’avancée de notre programme avec Orange Business, nous avons décidé de commencer à développer le reste de la solution. J’ai donc du me familiariser avec l’outils et pour accélérer le processus, j’ai décidé de télécharger la documentation officielle de Microsoft Power Automate et l’utiliser comme source de knowledge d’un GPT. J’ai donc commencé à conceptualiser l’architecture de la solution que j’ai divisé en plusieurs briques, qui deviendrait des flows Power Automate. Nous avions déjà la première brique de réception et d’enregistrement des PDFs dans un sharepoint. La brique suivante allait donc tourner autour de la thématique de traitement : étant donné que les analyses financières possèdent plusieurs composantes (paragraphes, graphiques détaillés, mise en page spécifique, etc…) il était primordial de s’assurer que nous les données soient mises dans un format optimal avant d’être fournies à un LLM. Afin minimiser le risque d’hallucination, il m’a semblé pertinent d’utiliser un OCR afin de ne se servir uniquement des éléments textuels des analyses financières. Je suis parti du principe que si une information est primordiale pour un analyste financier, elle se doit d’être écrite (ex : graphique détaillé -> explication du graphique). Ainsi j’ai commencé à me pencher sur une brique à travers laquelle les pièces jointes reçues passeraient par un OCR pour créer une version .txt de chacune des pièces jointes. Néanmoins ce pipeline ne vient pas sans complication : comment différencier les pièces jointes étant passées dans l’OCR de celle qui viennent d’être enregistrées et qui n’ont pas encore leurs versions .txt ? J’ai donc créé une nouvelle colonne au sein du sharepoint dans lequel les pièces jointes sont enregistrées afin de m’en servir comme un flag dans power automate : les pièces jointes ayant leurs versions .txt serait qualifiées « d’analysée » et inversement. Ensuite, j’ai dû intégrer l’appel LLM dans la boucle, il s’agit donc d’un GPT Microsoft, fondé sur le modèle GPT4o d’OpenAI.

Jusqu’ici, notre solution est capable de :

* + D’enregistrer les pièces jointes reçues par email dans un sharepoint.
  + De détecter lorsqu’une nouvelle pièce jointe est ajoutée au sharepoint pour en créer une version .txt.
  + Identifier les versions .txt qui aient déjà été synthétisée.
  + Combiner les synthèses en un regroupement de synthèse (et non une synthèse de synthèse pour éviter toute perte d’information).
  + Appliquer un format HTML à ce regroupement.
  + Envoyer le regroupement par email.
  + Mettre à jour les métadonnées des versions .txt étant été analysées.