Useful?	Type of Lit	Database	Title (+ hyperlink)	Summary	Why could it be useful?	Idea for our research contribution	Keywords	Member
Maybe	Academic	Google Scholar	Mortal Multi-Armed Bandits	Limited lifetime of ads so variant of k-armed bandits with a stochastic lifetime then expire. They outperform the standard MAB as they show less regret. Focused on the assumption that each arm exists perpetually. Mortal --> possible death of the relevance of ad. Takeway: regret has to be one our variables as it is how we evaluate the performane of each MAB.	Understanding how time can reshape the algorythm of MAB. The acticle was quite complex, maybe find a subject regarding time variability but make it easier: testing if variable arms or variable lifetime leads to the less regrets?		MAB, k-armed bandit	Pauline
Yes	Academic	Google Scholar	Bandit Algorithms Applied in Online Advertisement to Evaluate Click-Through Rates	Variations of UCB and best one to optimise CTR. Findings: UCB1Tuned is best --> By effectively balancing the exploration-exploitation trade-off, UCB1Tuned can enhance CTR & ad strategies	To look/analyse/use a specific MAB method: UCB for CTR		k-armed bandit, Upper confidence bound	Azahra
Yes	Academic	Google Scholar	Optimizing Click-Through Rates in Online Advertising Using Thompson Sampling	The advantage of TS sampling than other MAB methods --> able to integrate user behaviour data & ad contextual info to dynamically predict CTRs, even when there's sparse data or fluctuating ad performance	To look/analyse/use a specific MAB method: Thompson Sampling for CTR If we need sources to criticise A/B Testing Has a nice lit review!		multiple armed bandits, thomas sampling	Azahra
Maybe	Professional	Ayden	Optimizing payment conversion rates with contextual multi-armed bandits	Contextual MAB framework effectively optimized payment conversion rates by dynamically selecting the best strategies based on real-time payment contexts.	Good for understanding importance of context & conversion rate other than clicks No proper conclusion made + specific to payment as conversion (not clicks)		multiple armed bandits, conversion rate	Azahra
Maybe	Academic	Google Scholar	A Multiarmed Bandit Approach for House Ads Recommendations	Shows MAB effective in CTR and add-to-cart rates. Considers over time ads become less effective (ad fatigue) & assessed using bundle of ads	Considers non-stationary rewards & highlights personalisation. Could be too much within 'deep neural networ' for personalisation		multiple armed bandits, click through rate	Azahra
Maybe	Academic	Google Scholar	Showing Relevant Ads via Context Multi-Armed Bandits	Their CMAB balances exploration & exploitation by leveraging the structure of user query and ad spaces --> Lipchitz: similar queries paired with similar ads yield similar payoffs, so the algorithm can from observed data to unseen contexts	TBH depends on the data we have (yahoo)		multiple armed bandits, click through rate	Azahra
Maybe	Academic	Google Scholar	Pure Exploitation in Finitely-Armed and Continuous-armed bandits	Long article about the exploitation vs exploration trade-off, basically how long should you spend looking for resources (exploration) regarding the amount of value it is grogoing to bring, in this case, how much regret is lowered (exploitation). Cumulative regret can be minimized only if simple regret is minimized.	understanding exploitation and exploration trade-off		Exploration vs exploitation	Pauline
Yes	Academic	Google Scholar	Batched Multi-Armed Bandits Problem	Mix between batch learning and online learning. Experiment with static grid and adaptive grid. To achive the optimal minimax regret, it is necessary that M (number of batches) is within logaruthmic factors.	The regret analysis for batched stochastic multi-armed bandits remains underexplored: maybe find grey aea here		Batches	Pauline
Yes	Academic	Google Scholar	Top Arm Identification in Multi-Armed Bandits with Batch Arm Pulls	Twitter --> find the users that tweet the most about a topic. identify top k-armpes by looking at the total number of optimal batches. Aiming to reduce batch complexity by finding the amount of batches to correcctly identify the top k-arm. Fixed confidence setting --> number of k-arms with fewest batches possible, fixed budget --> given a batche number goal is to identofy top k-arms	Following the acticle above, this one was cited and I liked the real life application similar to ours. Maybe can trry to test many batches sizes and try to find an optimal one (without finding THE one). Also can look at regret but also at optimality gap to compare different performances.		Stochastic convex optimization, batched optimization, parallel computing	Pauline
Yes	Academic	Google Scholar	Online Interactive Collaborative Filtering Using Multi-Armed Bandit with Dependent Arms	Online interactive collaborative filtering approach using a MAB model with dependent arms. Traditional recommender systems face challenges such as the cold-start problem and lack of contextual data. This paper introduces a generative topic model that clusters dependent items (arms) to improve reward predictions. The model leverages particle learning for efficient online inference and integrates with bandit selection strategies like Thompson Sampling and UCB. Empirical evaluations on movie and news recommendation datasets.	UCB & Thompson sampling used + cluster parameters + coding for Interactive Collaborative Topic Regression (ICTR) mode (not sure if that is gonna be usefull + based on r 2 popular realworld dataset: Yahoo! Today News and MovieLens	Consider the time-varying property in user preferences for better online recommendation + provide a comprehensive regret analysis	Cold-start problem, collaborative filtering, MAB, Topic modeling, particle learning, matrix factorization	Valentine
Maybe	Academic	Google scholar	Multi-Armed Bandits in Recommendation Systems: A survey of the state-of-the-art and future directions	Lit review of 1327 articles published from 2000 to 2020	To have deeper understanding of the theme and find areas to reserach further for our contribution like cold-start problem,			Valentine
Maybe	Academic	Google scholar	Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit	Tracks changes in user preferences using a random walk process, helping to improve recommendation accuracy. It integrates with bandit selection strategies like LinUCB and Thompson Sampling. Experimental results on online advertising (KDD Cup 2012 dataset) and news recommendation (Yahoo! News dataset) show that this approach effectively captures contextual changes and enhances click-through rates (CTR).			Contextual MAB Time-Varying Reward Function Particle Learning Personalized Recommendation CTR optimization	Valentine
Yes	Academic	Google scholar	Optimizing Click-Through Rates in Online Advertising Using Thompson Sampling	Cites the article above -> more recent 2025!! Explores Thompson Sampling (TS) for optimizing click-through rates (CTR) in online advertising, addressing the challenges of dynamic user behavior and changing ad pools. By utilizing Bayesian inference, TS continuously updates the probability distribution of ad click rates, allowing advertisers to make real-time ad selection decisions even with sparse or missing data. Experimental results demonstrate that TS outperforms traditional A/B testing and other bandit algorithms in maintaining high CTRs while efficiently balancing exploration and exploitation.	I think this one is going to be super important for us / main one we will base ourselves on. Emphasis on: real-time decision making, handling spare data, balance exploration and exploitation (learn and earn), Bayesian approach, CTR as optimization goal.	Investigating how TS can handle multiple concurrent product recommendations rather than selecting a single ad at a time. Adapting the algorithm to multi-slot recommendations, where multiple products must be shown in a single interaction (e.g., e-commerce homepage). Integrating user demographics, browsing history, and contextual signals (time of day, device type, etc.) into TS-based recommendation models. Exploring how contextual bandits (LinTS, LinUCB) perform in product recommendation scenarios compared to standard TS. Examining how TS can ensure diversity in product recommendations while still optimizing for engagement. Benchmarking Thompson Sampling against other MAB methods (UCB, ε-Greedy, Exp3) in real-world product recommendation datasets. Testing in different domains (e.g., fashion, electronics, groceries) to assess domain-specific variations in algorithm performance.	Thompson Sampling (TS) Multi-Armed Bandit (MAB) Click-Through Rate (CTR) Optimization Bayesian Inference Real-Time Ad Selection	Valentine
Maybe	Professional	Google scholar	The Nonstochastic Multiarmed Bandit Problem | SIAM Journal on Computing	Introduces an adversarial bandit framework, where rewards are controlled by an adversary rather than a stochastic process. The authors propose EXP3 (Exponential-weight algorithm for Exploration and Exploitation), a randomized algorithm that efficiently balances exploration and exploitation under worst-case scenarios, minimizing regret at a rate of O(√(KT ln K)). Their theoretical analysis establishes lower bounds on regret and extends their findings to game theory, showing that their approach can be used in repeated unknown games to achieve near-optimal performance.	Ref list of the other thesis work		Adversarial MAB EXP3 Algorithm Regret Minimization Exploration-Exploitation Trade-off Game Theory Applications	Valentine
Yes	Professional	Google Scholar	Using Confidence Bounds for Exploitation-Exploration Trade-offs	Presents an upper confidence bound (UCB) algorithm that achieves improved regret bounds for both the adversarial bandit problem with shifting and associative reinforcement learning with linear value functions. The findings demonstrate that confidence-based decision-making can significantly enhance learning efficiency in uncertain environments, making it a powerful technique for online learning and adaptive decision-making.	Reducing regret, shifting consumer preferences (adversial bandits with shifts)	Explore hybrid approach: combining TS with UCB for better performance in diverse recomendation settings. Adapting the adversarial bandit shifting model to handle seasonal trends and market-driven fluctuations in product popularity.	Confidence bounds, UCB algorithm, reinforcement learning with linear value functions	Valentine

Yes	Academic	Google Scholar	Facing the cold start problem in recommender systems - ScienceDirect	This artcile describes the cold-start problem in recommendation systems. Proposes a system that providees predictions for new users, a mechanism that takes into consideration their demographic data and finds "neighbours" --> more possibilty to have a similar preference. Proposed system performs better in casees where a large number of users are already registerd in the system (makes it easier to allocate the new user to a group of people their preferences align with) --> increased accuracy of ratings prediction. System works as follows: new user is allocated to a group, and a rating prediction mechanism derives ratings for items, then the ratings are weighted out.		Which MAB algorithm is most effective in mitigating	Cold-start problem, content-based models, collaborative filtering	Esther
Yes	Professional	Google Scholar	[0805.3415] On Upper-Confidence Bound Policies for Non-Stationary Bandit Problems	Changing environment (at unknown time instants) --> establishes that UCB policies can also be succussfully adapted to non-stationairy environments. Analyses two algorithms: the discounted UCB and the			Non-stationary bandits, UCB, Reinforcement learning, deviation inequalities	Esther
Maybe	Professional	Google Scholar	Hierarchical Bayesian Bandits	Studies hierarchical Bayesian bandits, which is when different arms of the bandit share information --> different target segments share data. (If one segment has liittle data (cold-start problem), it borrows informationi from another segement. Proposed a natural hierarchical thompson samplng algorithm.	Understanding the hierarchical bayesian bandits and how they can be used for the cold-start problem.	How could an MAB model make use of hierarchical Bayesian bandits to solve the cold-start problem? / There are suggestions in the article, but they seem a little too complex for a bachelor tthesis.	Hierarchical Bayesian bandits, upper bound, lower bound, regret	Esther
Maybe	Professional	Google Scholar	https://dl.acm.org/doi/abs/10.1145/1557019.1557161?casa_token=f3XNeYHcVg0AAAAA:82e8-ZBOAOdD6jELZ0SaoCSaiKBx001DrSCiKpnPBL5I7cpEvTj4t7Yeb1iz4A1HYd6Hf9WkFkVR	Explores bounce rates in the context of sponsored search advirtisement. The paper proves that bounce rates are an effeectve measure of user satisfaction. Asks the question: Can we predict bounce rate by analysing the features of the advirtisement? If so, advertisers and search engines could more accurately predict effectiveness of their ads before using them.	Defining bounce rates, seeing how bounce rates can be predicted	Which MAB model predicts bounce rates most accurately?	Bounce rate, sponsored search, machine learning	Esther
Yes	Professional	Google Scholar	Reinforcement learning: exploration–exploitation dilemma in multi-agent foraging task | OPSEARCH	Uses a variety of models / learning policies on multi agent foraging task (specfic) in terms of the exploration-exploitation "dilemma".	Defining exploration-exploitation		Exploration - exploitation dilemma, reinforcement learning	Esther