Section: Psychiatry

Review Article

ROLE OF ARTIFICIAL INTELLIGENCE IN TOBACCO CONTROL: A SCOPING REVIEW

Kritika Sisodia[1], Shreya Gupta[2], C.S. Sushil[3], Vinod Daria[4]

[1]*Assistant Professor, Department of Psychiatry, Government Medical College, Bundi, Rajasthan, India.*
[2]*Senior Resident, RNT Medical College, Udaipur, Rajasthan, India,*
[3]*Senior Professor and Principal, Bharat Vikas Parishad Hospital, Kota, Rajasthan, India*
[4]*Professor, Department of Psychiatry, Government Medical College, Kota, Rajasthan, India*

Received : *22/02/2026*
Received in revised form : *07/04/2026*
Accepted : *28/04/2026*

Corresponding Author:
Dr. Kritika Sisodia
Assistant Professor, Department of Psychiatry, Government Medical College, Bundi, Rajasthan, India.
Email: kritikasisodia25011994@gmail.com

DOI:10.70034/ijmedph.2026.2.225

Source of Support: Nil,
Conflict of Interest: Nonedeclared

Int J Med Pub Health
2026; 16 (2); 1329-1351

ABSTRACT

Background: In the recent years, artificial intelligence (AI) is marking significant impact in fields such as psychology, medicine, public health, etc. and researchers are using AI tools to tackle complex issues in tobacco research. In this scoping review, we aimed to summarise the published evidences on role of AI interventions in relation to tobacco cessation, motivation, relapse, cravings and identification of promoting factors.

Materials and Methods: For this scoping review, we employed the PCC (Population, Concept & Context) framework as per the PRISMA-ScR guidelines. We explored literature using the following four databases- PubMed, Google Scholar, The Cochrane Review & DOAJ. Articles were included if they were peer-reviewed, published in English between 2019 & 2024 and reported the use of AI techniques like conversational agents(CA), machine learning(ML), natural language processing(NLP) etc. in tobacco use. We excluded studies done on individuals using vaping or e-cigarettes. The extracted data included title, author, year of publication, type of study, sample size, socio-demographic characteristics, study strategy, type, name, frequency & limitation of AI intervention and outcome of study.

Results: The initial search yielded 1790 articles which were uploaded to Rayyan. After removing duplicates (n=83), the remaining articles were screened and the relevant 42 articles were finalized among which 1 was narrative review and 4 were systematic review and meta-analyses. The studies were diverse in design, strategy, type and platform of AI used. Chatbots (rule based and embodied) and ML were used in some studies to promote abstinence and increase motivation. Some used hybrid AI model for predicting smoking status, lapse risk and treatment outcome while others were iterative developmental study in which they developed algorithm or model of AI and then tested it.

Conclusion: While AI hold considerable potential for supporting tobacco cessation, most existing applications relied on rule-based design that provide limited adaptability and resemble standard text-messaging programs. Some studies implemented AI on small sample which limits their generalizability while in some studies loss to follow up was high. Given the high prevalence of tobacco use in adolescents, we found only 2 such studies. This gap highlights the necessity for further well designed studies employing advanced AI driven models.

Keywords: Tobacco, Nicotine, Smoking, Cigarette, Artificial Intelligence, Machine Learning, Chatbots, Deep learning.

INTRODUCTION

India holds 2nd position after China in both consumption and production of tobacco.[1] The WHO's GATS-2 survey, conducted in 2016-2017 states that 28.6% of population in India consumes tobacco with current cigarette & bidi smokers being

4% & 7.7% respectively. Despite this, merely half of the consumers (55.4% current smokers and 49.6% consumers of smokeless tobacco products) are planning to quit.[2]

Interventions for Tobacco Use: Conservative approaches rely on manual, laborious & self-reported epidemiological surveys, regression-based prediction, NRT/ pharmacotherapy, counselling, biochemical verification, etc. Lately, researchers have been using machine learning and other AI methods to study tobacco in new ways. They look at how genes and psychology,[3]may make someone more likely to get addicted; track habits and behaviours of people who use tobacco; create apps and digital tools to help people quit; and identifying for whom treatment is most effective.[4]

Concept of Artificial Intelligence: AI is the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making.[5] [Figure1]

ML is a discipline at the intersection of data science and artificial intelligence with a focus on building algorithms that enable systems to learn from data and make predictions or decisions without explicit programming. Supervised algorithms deal with clearly labelled data and typically serve 2 purposes: classification or regression[6]. Unsupervised learning algorithm inspects the data to detect trends, uncover correlations & links by clustering, frequent pattern detection, and dimensionality reduction.[6]

Generative AI models act like students learning from large books of examples. Most popular types are:

- Large Language and Small Language Models (LLMs & SLMs): LLMs train on large text for recognizing and generating texts. Two well-known large language models are the Generative Pre-trained Transformer (GPT), used in ChatGPT and Bidirectional Encoder Representations from Transformers (BERT).[7]
- Generative Adversarial networks (GANs): used for changing images in various ways (from style to color or content) or synthetic data generation (for training other models).[8]

Deep Learning models (ANNs, CNNs, RNNs, LSTMs) are advanced forms of ML that use neural networks with many layers, ideal for tasks involving big data and complex pattern recognition like voice recognition systems, image classification, etc.[9]

Natural Language Processing (NLPs) models focus on interaction between computers and human language. These enable machines to understand, interpret, and generate human language. NLPs are applied to train chatbots, sentiments analysis, translation of a language, etc.[10]

Conversational Agents are programs used to assist healthcare services because they are accessible, can engage users in human-like conversations; most common type being chatbots.[11]

Hybrid AI models combine different AI techniques to improve performance and address complex challenges that single-method approaches might struggle to solve.[12]

Study Intentions: Given the future possibilities and infancy of use of AI tools for tobacco related disorders, there is a clear need to systematically condense the available data. In this scoping review, we tried to answer the following review questions:

1. What are the published evidences on role of AI intervention in relation to tobacco cessation, motivation, relapse, cravings and identification of promotion & environments?
2. What are the benefits & shortcomings of AI technologies used in the above mentioned context?

METHODOLOGY

This review was executed as per the PRISMA-ScR (Preferred Reporting Items for Systematic Review and Meta- Analyses Extension for Scoping Reviews) guidelines and the protocol is registered in Protocols.io (Protocol Integer ID: 227101). We employed the PCC framework recommended by Joanna Briggs Institute (JBI).[13]

1) Population(P): Individuals using tobacco products or those exposed to tobacco related content or environments.
2) Concept(C): Role of AI methods in tobacco cessation, motivation, relapse, cravings and identification of promotion & environments.
3) Context(C): Any domain (clinical, community, online, public health records or research) with no restriction on geographic region, cultural background or study design.

Search Strategy

Data Sources: We explored peer-reviewed literature using the following four databases- PubMed, Google Scholar, The Cochrane Review & DOAJ (Directory of Open Access Journal) published between 2019 and 2024 on 4th April, 2025. An additional search to identify relevant literature was also done through bibliography of included articles. Last searches were completed on 22ndJuly,2025. The detailed search strategy has been explained in Appendix-1

Inclusion/ Exclusion Criteria-

Articles were included if they were peer-reviewed, published in English between 2019& 2024 and if they:

- Addressed individuals who use tobacco products or encountered tobacco related content (e.g. human texts, blogs, images, videos, marketing schemes, and social media content) or environments.
- Reported the use of AI techniques like CA, ML, NLP, etc.
- Conducted in any setting like clinical, community-based, online/ digital, public health records or research.

- Had any type of study designs like RCTs, developmental studies, observational studies, etc.

We excluded studies addressing multiple substance abuse, vaping or e-cigarettes, molecular/genetic level research and having comorbid mental illness.

Study Selection: Studies (N=1790) identified from databases & citation tracking were imported to Rayyan. After removing duplicates (n=83), studies were screened based on title and abstract. The next step that we did was to assess the full text of articles that met the inclusion/exclusion criteria. The two researchers (KS and SG) met to resolve the disagreements. Final studies included in the review were 42. [Figure-2: PRISMA flowchart]

Data extraction: Two authors (KS and SG) extracted the data from final included records. The extracted data included title, author, year of publication, type of study, sample size, socio-demographic characteristics, study strategy, type, name, frequency &limitation of AI intervention and outcome of study (Appendix 2 & 3).

Integrated Findings

The included studies' characteristics can be found in Appendix 2 & 3 and detailed description of it is mentioned below:

Hallmarks of Studies: Majority of the articles (10) included were published in the year 2024. Year-wise distribution of remaining articles are as follows: 6 in 2019, 9 in 2020, 4 in 2021, 5 in 2022,and 8 in 2023.

Our review included various types of studies- 8 RCT, 5Iterative developmental, 6 Mixed-Method, 3 Descriptive, and Analytical Cross-Sectional, 3 Cohort (Prospective, and Retrospective, 9 Secondary data analysis, 1 Qualitative, and 2 Retrospective Observational studies. Our review also contains 1 narrative review, 1 scoping review, and 3 systematic review& meta-analysis.[14-30]

The sample characteristics reflected considerable diversity. 29 articles had human participants (Total N=4, 48,168) ranging from 14, to 3,43,481,participants. A key feature of our review was the inclusion of studies with non -human participants like community posts, subset of messages, tweets, advertising videos, smoking images, subreddits and utterances by volunteers. Two studies featured entries in EMRs as sample.[31-49]

Most studies recruited participants over the age of 18 years, but the inclusion of adolescents in two studies,[40,41]represented a significant strength.

Out of 8 RCTs, 6 had active control group: usual clinical/pharmacotherapy,[14,15]and smart-phone app.[16-20]2 studies tested different chatbot modalities as a part of their intervention groups.[19,27] Participant loss to follow up ranged from 14%,[16]to 80%,[18]in intervention groups, and 14.6%,[16]to 90%.[18]in control groups.

Summary of Ai algorithms used: We incorporated studies employing different types of AI approaches, encompassing ML, CA, NLP, etc. [Table 1]

Various platforms of AI intervention were app based, web based or software/ developer platforms. 8 studies used app-based,[14-18,23,37,43,47]platforms like telegram,[14] So-Lo-Mo app,[15] Smoke-Free app,[18]etc. 10 studies employed web-based platforms,[19-29]like Vet Flexiquit,[20] Google Firebase,[32]etc. Softwares,[24-45]like RASA,[24] Tensor Flow,[38] WEKA,[45] etc. were used in building and testing AI tools across 6 studies.

RESULTS

We analysed a range of AI tools for addressing different aspects of tobacco use. 12 studies,[4,17,33,34,36,39,40-46]employed ML algorithms to predict smoking status, lapse, cessation, etc. Leanne Nicole Siegel,[42] Mona Issabakhsh,[36]and Kavita Rijhwani,[46] predicted variables for quit status through different models of ML. Leanne used SML model (LR) to highlight 3 quitStart app features- "I slipped button"-less likely to quit (negatively related), number of times they viewed "Tips" pages in app and no. of badges they completed (both were positively related). Mona Issabakhsh,[36]analyzed PATH survey variables using 4 ML algorithms: RF, GBM, GLM, XGBoost. Top variables of cessation shown by GBM & RF (both performed best) were: age of initiation, past 30 days polytobacco use, etc. TreeSHAP determined effect direction of each variable and it showed that more e-cigarette use, less cigarette use in past 30 days, higher BMI & smoking initiation at older age were more likely to quit. Kavita Rijhwani,[46]compared ML algorithms (NB, SMO, RF, J48, Decision stump) for quit status. Decision stump & SMO- best prediction & accuracy, Naïve Bayes (0.05s) & Decision stump (0.05s) took least time to build a model, Decision stump (55.87%) & SMO (52.21%) had most correctly classified instances.

Elastic Net Cox Regression tools were used by Emily T,[43] and R. Schuting,[44] to identify predictors of lapse. Emily T.[43]found the top 3 predictors: Self-reported odds, Motivation & Confidence to avoiding smoking which they incorporated in Smart-T2 app to make JITAI effective. ML models used by R. Schuting,[44]retained 17 out of 26 candidate predictors of lapse from 2065 EMAs & the approximated model retained 6 predictors- feeling irritable, having cigarette available, being around as smoker, being in an area where smoking is discouraged, permitted and alcohol consumed in past hour.Two studies,[40,41]implemented GUIDE regression tress & SMOTE with RF classifier to predict smoking status among adolescent girls. Nayoung Kim,[40]found that exposure to tobacco promotions, support for outdoor & indoor SFPs were predictors of susceptibility & current tobacco use from GYTS (2013 & 2017) data. On the other hand, Sara V. et al,[41]found that SMOTE with RF classifier was best to identify predictors which came as girls' independence &connectivity, social environment and peer-influence.Pranta Roy,[34]assessed impact of smoking on academic life which was tested on different ML classifiers. On comparison they found that best performing in accuracy were KNN & RF (80%), in precision was SVM (87%) and in recall was RF (81%). Jamie Faro,[17] in their study found that quit rate among participants was more for ML recommender text messages than standard text messages. Noberto Francisco,[33]conducted study among tobacco cessation requesting users of Mexico and was able to correctly identify 78.6% by RF. Similarly, to detect smoking status of online cessation community users through posts and blog comments, Xi Wang,[39]revealed that overall best performing model was AdaBoost using J-48 decision tree (F1-0.759, AUC-0.837). Lara N.[4]applied CART to candidate variables of two cohorts (TC & VC) and the resulting decision trees showed that delay discounting task was strongest & most consistent predictor of smoking cessation treatment outcome.Apart from ML, CAs emerged as a prominent tool for tobacco control. The level of intelligence of most chatbots was rule based with predefined scripts and only 3 studies,[20.21,29]incorporated Embodied chatbots i.e. that had ability to understand natural language & maintain conversation flow. Mode of interaction of CAs was via text but 3 studies had chatbots communicating through text with emojis,[47] text with speech,[29] or voice of avatar with hand and mouth gestures.[29] Most chatbots maintained two-way conversational flow except Katsunori Masaki,[16](automated, not fully bi-directional) and Alice Alphonse,[47] (one-way). Laura Carascco,[15]et al tested mobile app which was connected with an AI system designed to learn from patient interests through their interactions with the app. Other characteristics of chatbots are enlisted in [Table 2].

Our review identified six studies that incorporated rule-based chatbot delivery system in combination with ML algorithms. This hybrid design was intended to increase the chatbot's flexibility and create more human-like responses.

Eduardo Olano Espinosa,[14]checked effectiveness of chatbot for smoking cessation using Bayesian probabilistic decision process (ML) which acted as the "brain" interpreting user messages. Treatment including a chatbot was more effective than usual clinical practices as statistically significant results were obtained for continuous abstinence, mean interaction time & intensive use. Fahad Almusharraf,[22]used supervised ML to train chatbot's natural language understanding (NLU) **system**s so it could better recognize pros and cons of smoking from smokers' text inputs and provide more accurate reflection in motivational interviewing for ambivalent smokers.

Andrew Brown,[25]made 4 chatbot versions by incorporating deep learning ML and NLP in chatbots. MIBot v4.7 asked questions without reflections. With each upgraded version of MIbot till MIbot v5.2, level of MI style conversation reflections increased. MIBot v5.2 is the only version associated with significant increase in importance to quit smoking($p<.001$) and readiness($p=.01$).

J.B. Bricker,[23] conducted 11 step developmental study. 11th step yielded first LLM based quit smoking program (initially step1 to 10 conversations with rule based chatbot) which enabled users to ask open ended questions about quitting. Ahson Saiyed,[24]developed the Technology Assisted Motivational Interviewing Coach (TAMI) for tobacco cessation, a digital CA that used ML models to deliver MI and create personalized quit plans. For this, the chatbot performed 3 key tasks: onboarding (for initial user assessment), MI (for users not ready to quit) & outboarding (develops quit plan). Tavleen Singh,[30]developed an innovative CA for just-in-time cravings' intervention. The chatbot built 5 cravings-related scenarios describing users' needs. Using supervised ML (JITAI) & deep learning models, the chatbot was made more adaptive & human-like.

Mathew M,[35] in their study explored the fact that objects and settings found in images of daily life can be used to identify environments associated with smoking. For this they employed hybrid AI method comprising deep learning and Supervised ML. The deep CNN acted as a 'feature' extractor from each photo (e.g. Patio, pool table etc) and then LoR calculated the probability that CNNs' detected objects/settings were smoking env or not.Daniel R. Harris,[49]used a different approach using problem list entries in EHRs. They processed 1667 entries associated with tobacco use employing hybrid AI techniques that combined ML based CLAMP-NLP and rule-based AI models. CLAMP models were trained on tobacco-related sentences so that they can classify problem entries into current user, former-tobacco & non-tobacco. And if ML-CLAMP sometimes made mistakes, then rule based AI stepped in to check the text for specific keywords/ patterns & corrected the obvious errors. Likewise, Andrea Caccamisi et al,[45] explored Swedish EMR data and applied NLP & ML to develop 32 different predictive models enabling automatic classification of unstructured information on smoking status of patients. Pre-processing by NLP led to changes in sentence frequency, classifier type, tokenization & attribute selection followed by development of ML algorithms (SVM, KNN, NB, J48). Ultimately, the system produced classified smoking status (current smoker, ex-smoker, non-smoker, unknown).

Cheng Chein Li,[38]compared different ML & DL models (SVM, RF, LoR, KNN, CART, NB & ANN) to construct predictive models for smoking cessation. Out of all, ANN performed slightly better with a sensitivity of 0.704, specificity of 0.567, accuracy of 0.640 & ROC value of 0.660.

Khishigsuren Davagdorj,[48]dealt with Class imbalance problem in ML for the prediction of smoking cessation. Class imbalance means one outcome dominates the dataset, causing models to ignore the rare but important cases. First, he used Lasso for excluding irrelevant features from socio-demographic data for smoking cessation-17 out of 22 included. Then used multicollinearity analysis which confirmed no correlation among features. Balanced data was generated via oversampling technique-SMOTE and ADASYN. Classifiers like GBT and RF etc. were applied to construct the prediction models for successful quitting. Comprehensive analysis demonstrated that Gradient Boosting Trees (GBT), Random Forest (RF) and multilayer perceptron neural network (MLP) classifiers achieved the best performances in all subjects and each gender when SMOTE and ADASYN were utilized. The review-level studies provided complementary perspectives, adding worth to our scoping review [Table-3].

Identified Shortcomings of Included Ai Methodologies: AI applications in tobacco research still face several limitations & ethical concerns which are insufficiently addressed in studies included. Most studies provided limited time and data to adequately train ML models resulting in reduced accuracy, misclassification, limited generalizability and thus affected the reliability of their findings. They often work on down-sampling or over-sampling techniques which results in removal of relevant data. The complex models like deep neural networks often lack transparency and are labelled as "black-box" models. This makes it difficult for clinicians to fully understand exactly how input variables influence the output.

Chatbots in most of the studies were rule-based and did not fully capture human complexity, especially emotions, changing motivations or deep psychological processes. These could not process open-ended input or adapt dynamically like a human or an advanced chatbot leading to fixed and repetitive responses which led to declining engagement over time. The engagement with the chatbot depends on the digital literacy of the user and anthropomorphism of the chatbot. A very interesting finding was reported by Fahad Almusharraf that chatbot implemented running head start technique even when smoker was motivated which was frustrating for users. Some studies also reported that chatbot was not able to detect best time to send message to user which was very annoying for users. Although chatbots and virtual assistants offer scalable, round the clock support for tobacco cessation, but further research is needed to improve their quality.

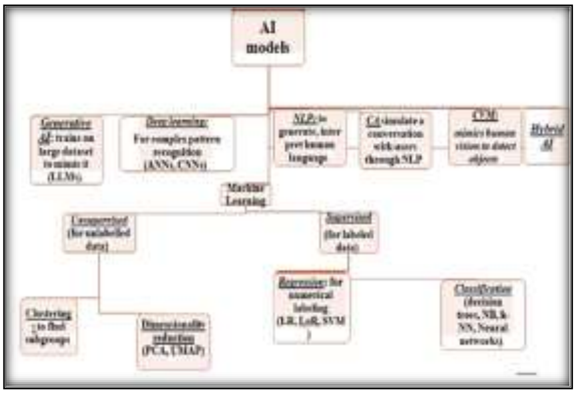

Figure-1: Various AI models

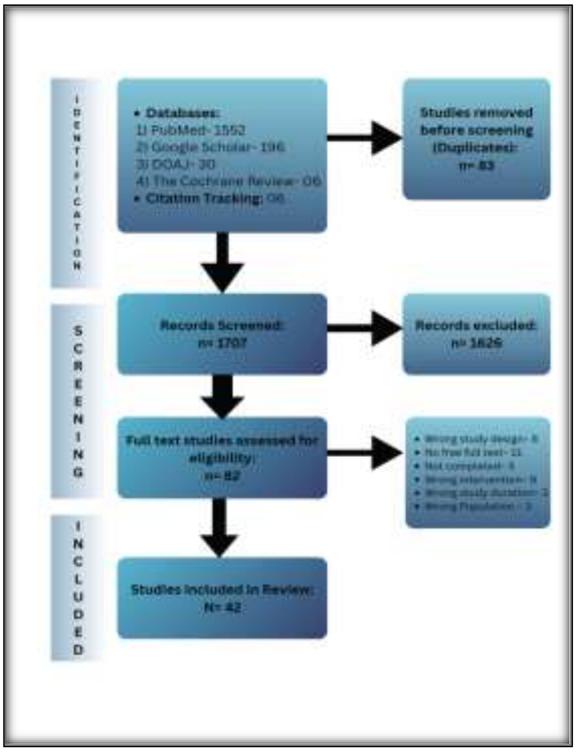

Figure 2: PRISMA Flowchart

Table 1: Summary of AI framework

S No.	AI METHOD	Studies
1)	1 supervised ML model	Lara N (CART), Noberto Fransisco& Sara. V (RF), Jamie Faro (ML recommender system), Nayoung Kim (GUIDE), Leanne Nicol Seagle(LR)
2)	>1 supervised ML model	Pranta Roy, Xi Whang, Mona Issabaksh, Robert Schuting, Kavita Rijhwani,Emily(JITAI+Elastic net regression)
3)	Deep Learning model	HuseiynKuchukali(LLM), Robert Lakatos
4)	Rule based CA	LinweHe(MI Vs CC),Katsunori Masaki, Alice Alphonse, Hollie Bendoti, Lotte Lewis, Olga Perski, LinweHe(only MI), Ayaka Kyato
5)	Embodied CA	Kate Loveys, Jamie L Heffner, Maria Karelka
6)	Hybrid AI	Eduardo Olano, Fahad Almushraf, Andrew Brown, Matthew M, JB Bricker, Ahson Saiyed, Tavleen Singh, D.R Harris
7)	AI driven smartphone app	Laura Carrasco
8)	Supervised ML with NLP/ DL	Andrea Caccamisi,Davagdroj and Cheng Chein li

Table 2: Characteristics of included Conversational Agents

Studies	Platform	Content delivered by chatbot	Engagement	User satisfaction	Accessibility	Limitations of chatbot
Linwe He(MI vs CC)	Web interface	Mi style vs CC style	Better for MI style	Better for MI style	2 sessions	Declining engagement
Katsunori Masaki	CASC smartphone app	Interactive counselling	Higher engagement in successful quitters	NA	Several times/day upto 24 weeks	Rule based chatbot
Olga Perski	SmokeFree smartphone app	BCTs(positive reinforcement& motivation)	Better for SmokeFree app having chatbot	NA	twice daily via notifications + on demand support via text	every app open was counted as a new login
Alice Alphonse	Smoking cessation app	Motivational messages of Anthropomorphism quality	High at the time of strong craving	High due to anthropomorphic experience.	Twice daily check ins in 1st 2 weeks, reduce and stop at 90 days	Check-ins perceived as annoying when check-ins during working hours.
Linwei He (only MI)	Web interface	MI style v/s neutral style	No difference in the two styles	NA	2 sessions	Very slight difference between both styles of chatbot
Lotte Lewis	Web interface	MI style v/s CC style	No significant diff in two styles.	No significant effect between 2 styles of chatbot	One 8 minute conversation	Used little NLP limiting the quality of interaction
Ayaka Kyoto	Ascure smoking cessation app	Motivational messages+ Personalised advice for symptoms	High engagement due to interactive sessions	NA	6online sessions, multiple interactions	Reactive sessions not proactive(messages delivered when press call button)
Hollie Bendoti (Focus)	Google Firebase	Personalised quit plan, 'moments of crisis' support with educational & motivational content	Overall positive engagement	Positive satisfaction	multiple times over 1-2 weeks(mean duration=58 mins)	Recommendations by chatbot are limited in generalizability
Kate Loveys	WHO publicly accessible website	Personalised quit plan + WHO recommended cessation strategies	Positive engagement	Overall good	NA	Some felt interaction robotic, not human like voice
Jamie L Heffner	Vet Flexiquit web based	Acceptance& Commitment therapy	Overall low engagement	100% satisfaction for those who engaged	6sessions;each of 10-25min	Time constraints, and rule based
Maria Keralka	Website	Acceptance& Commitment therapy	Highly positive	Highly positive	6sessions; each approx. 25 mins	Fixed responses, could not process open-ended input or adapt dynamically like a human or an advanced chatbot.
Laura Carassco-Hernandez et al	So-Lo-Mo app	Motivational messages	Engagement was **higher in participants who successfully quit smoking**.	High overall	Maximum 150 motivational messages could be received during 1 year study	AI was not able to detect best time to send best message to user. Only 150 messages could be sent.

Table 3: Overview of included review-level studies

Author	Total Included Sample	Findings
R. Vilardaga	33 apps	Majority of apps for cessation used psycho-education & self-tracking. Less commonly used features were personalised feedback, reward system, geo-location, just-in-time sensors & ML
Linwe He MSc	13 studies	Overall high acceptability & high engagement of CAs but potential publication bias was observed, bias due to self-reported abstinence
Hollie Bendotti	5 RCTs	Chatbots for cessation were either embedded in smartphone apps or based on social media or web interface and they significantly improved the rate of cessation at 6-month follow-up. Loss-to-follow-up & risk of bias were overall high.
Warren K. Bickel	12 studies	Studies revealed that ML algorithms like RF, classification trees, regression trees, etc. had immense potential to identify predictors of cessation and best predictor came out to be 'delay-discounting'. Overall, studies varied in quality with only a few having high methodological standards & relied heavily on self-reported measures.
R. Whittaker	10 studies	Chatbot intervention varied across studies (embodied, used social media platform or were part of smartphone app) and mostly positive outcomes for cessation was noted. Overall quality of studies was low, with methodological issues & low follow-up rates.

DISCUSSION

The objective of our scoping review was to highlight the extent of available evidence regarding the role of Artificial Intelligence in various aspects related to tobacco. We explored the utility of various AI techniques like ML, chatbots, deep learning, etc. and subsequently provide a clear and robust account of the included studies.

It is evident from the included articles that in today's scenario AI is being used primarily in two domains: I) predictive modelling, where mostly ML algorithms are applied to large datasets in order to identify individuals at high risk of lapse or to predict smoking status or cessation treatment outcomes. II) cessation support, where CAs, mobile health platforms and hybrid AI methods combining ML/ CAs with expert NLP and DL, etc. appear particularly promising as they offer tailored and personalized content which helps in motivation, abstinence, cravings, moments of crises, etc.

A large proportion of the studies employed supervised ML techniques with traditional algorithms like RF, LoR, SVM, decision trees, NB, etc. to detect smoking status/ quit status of participants. Others used more advanced approaches, including gradient boosting methods (eg.XGBoost, GBM), ANNs, CNNs, RNNs and transformer based models like BERT to a wide variety of tasks like image & text classification, monitoring social media trends, identifying tobacco related promotions & environments, etc. In addition to this, tree-based models like GUIDE and CART were also utilized for their ability to detect complex interactions thus identifying susceptible tobacco users or current users or success rate of cessation interventions. Two unique studies incorporated JITAI interventions to extract nicotine cravings' related scenarios and identifying factors for relapse.CAs like chatbots & virtual assistants emerged to be particularly helpful in tobacco cessation. We retrieved 18 studies that represented various forms of unique CA systems. Majority of chatbots engaged in interactive, two-way conversation in text form with the users to provide motivational content. 5 studies integrated chatbots within a smartphone app like SmokeFree app, CASC app, etc. while the remaining ones were available via web interfaces or software platforms like RASA. Kate Loveys, Jamie L. Heffner and Maria Karelka worked with embodied CAs with additional features like speech, virtual human-like avatars, etc. Despite the heterogeneity in study designs and outcomes evaluated, all studies reported positive user satisfaction and overall high engagement. Participants valued some aspects of the agents like easy accessibility, personalized & non-judgmental support, user-friendly interface and that the chatbots provided accurate, evidence-based information.

Hybrid AI methods were used in 11 included studies. Almost half of the articles explored the integration of rule-based chatbots with ML methods which allowed for better, context-relevant responses and thus better engagement and user trust. One study by Matthew M. used deep CNN with ML to identify smoking related environments from images of daily life. Whereas some studies used DL with ML to either classify tobacco-related problem entries in EHRs or to construct predictive cessation models or to deal with class imbalance problem.Apart from these, we also included 5 review-level studies exploring tobacco cessation apps, chatbot interventions for cessation & ML algorithms to identify predictors of cessation. These provided us with a broader perspective on the topic and helped in conceptualizing our findings by highlighting the trends & gaps which were not apparent in the empirical studies.

An important fact that needs to be highlighted here is that we couldn't find any studies on unsupervised ML. This absence indicates the current focus of tobacco research on prediction and classification tasks predominantly, which align more appropriately with supervised methods. Nevertheless, unsupervised ML methods might be promising in uncovering hidden patterns in tobacco use and detecting newer, more complex trends among tobacco users.

Despite growing research, we noticed that several gaps remain. We found only 2 studies specifically on adolescents and this paucity needs further research as tobacco use in this population is increasing globally. Only a handful of studies reported qualitative feedback on AI approaches, limiting the deeper insight into users' experiences and perceptions. Although India is the tobacco tycoon of the world, no studies targeted smokeless tobacco users residing here which represents that predominantly studies are from developed and high-income countries. This clear geographic gap highlights the need for future research in LMICs.The main takeaway of our review is that there is scarcity of high-quality studies in this area. Most of the existing literature is based on studies that are preliminary, pilot-level trials, with small sample sizes, methodological errors, reliance on self-reported measures and limited follow-up periods. RCTs which are considered as the gold standard of research work were very few (only 8 RCTs). Loss-to-follow-up (LTFU) was considerably high in most studies thereby threatening the internal validity and generalizability of study findings. All these factors made it difficult to draw firm conclusions about the effectiveness, scalability and real-world implementations of AI based interventions in tobacco control.

Strenghts, Gaps & Opportunities Ahead: Through this Scoping Review, we have tried to provide valuable insights into how AI has been applied in tobacco research. This review offers several strengths. We analyzed a wide variety of studies (RCTs, Observational studies, Systematic Reviews, etc.), covering diverse populations (adolescents, young adults), methodologies and AI intervention approaches (ML, CA, DL, Hybrid AI, etc.) and this helped us to present a balanced picture of use of AI in this field. The use of a structured search strategy and transparent inclusion criteria further adds to the rigor of our work. To our knowledge, no prior review has focused on this topic in the Indian context, making our review a novel contribution to the field.

Alongside these strengths, some gaps and limitations must be acknowledged here. The reviewed studies were highly heterogeneous in design, interventions and purpose, so conducting a formal assessment of methodological quality or risk of bias would have been challenging and could have limited the inclusivity of the review. Our focus was more on mapping the evidence rather than critically appraising the quality of included studies and this decision aligns with the primary objective of a scoping review. Apart from that, we included articles from only selected databases using Mesh subheadings and in English language and this restriction may have resulted in the omission of relevant studies as terminologies can differ with author to author. Lastly, being a scoping review the findings were descriptive in nature with limited generalizability and effectiveness or causality of interventions could not be established.

CONCLUSION

In conclusion, future research should focus on more methodologically rigorous studies emphasizing real-world implementations of AI techniques in tobacco use. Research should be more transparent, capturing more qualitative user experiences. People from low and middle-income countries and under-represented populations like adolescents and smokeless tobacco users should be included in studies.

REFERENCES

1. Food and Agriculture Organization of the United Nations, Tobacco production, 2021, Our World in Data, accessed August 2023.
2. Tata Institute of Social Sciences (TISS), Mumbai and Ministry of Health and Family Welfare, Government of India, Global Adult Tobacco Survey GATS 2 India 2016-17
3. 3. Xu, Yi & Cao, Liyu & Zhao, Xinyi & Yao, Yinghao& Liu, Qiang & Zhang, Bin & Wang, Yan & Mao, Ying & Ma, Yunlong & Ma, Jennie & Payne, Thomas & Li, Ming & Li, Lanjuan. (2020). Prediction of Smoking Behavior From Single Nucleotide Polymorphisms With Machine Learning Approaches. Frontiers in Psychiatry. 11. 10.3389/fpsyt.2020.00416.
4. 4.Coughlin LN, Tegge AN, Sheffer CE, Bickel WK. A Machine-Learning Approach to Predicting Smoking Cessation Treatment Outcomes. Nicotine Tob Res. 2020 Mar 16;22(3):415-422. doi: 10.1093/ntr/nty259. PMID: 30508122; PMCID: PMC7297111.
5. 5.Sheikh, H., Prins, C., Schrijvers, E. (2023). Artificial Intelligence: Definition and Background. In: Mission AI. Research for Policy. Springer, Cham. https://doi.org/10.1007/978-3-031-21448-6_268.34566
6. 6.Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, Siebourg-Polster J, Steiert B, Zhang JD. An Introduction to Machine Learning. Clin Pharmacol Ther. 2020 Apr;107(4):871-885. doi: 10.1002/cpt.1796. Epub 2020 Mar 3. PMID: 32128792; PMCID: PMC7189875.
7. 7. Håkansson A, Phillips-Wren G. Generative AI and Large Language Models - Benefits, Drawbacks, Future and Recommendations. Procedia Computer Science. 2024;246:5458–68.
8. 8.Wenzel M. Generative Adversarial Networks and Other Generative Models. 2023 Jul 23. In: Colliot O, editor. Machine Learning for Brain Disorders [Internet]. New York, NY: Humana; 2023. Chapter 5. PMID: 37988513.
9. 9.Adeel A, Gogate M, Hussain A. Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments. Information Fusion. 2020 Jul 1;59:163-70.
10. 10.Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2023;82(3):3713-3744. doi: 10.1007/s11042-022-13428-4. Epub 2022 Jul 14. PMID: 35855771; PMCID: PMC9281254.
11. 11.Mariani MM, Hashemi N, Wirtz J. Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. Journal of Business Research [Internet]. 2023 Mar 21;161:113838.
12. 12.Zaheer F. Hybrid AI models: combining symbolic reasoning and deep learning for enhanced decision-making [Internet]. 2022
13. 13.Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. 2015 Sep;13(3):141-6. doi: 10.1097/XEB.0000000000000050. PMID: 26134548.
14. 14)Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, Matilla-Pardo B, Serrano Serrano ME, Martinez-Suberviola FJ, et al. Effectiveness of a Conversational Chatbot (Dejal@bot) for the Adult Population to Quit Smoking: Pragmatic, Multicenter, Controlled, Randomized Clinical Trial in Primary Care. JMIR mHealth and uHealth. 2022 Jun 27;10(6):e34273.
15. 15)Carrasco-Hernandez L, Jódar-Sánchez F, Núñez-Benjumea F, Moreno Conde J, Mesa González M, Civit-Balcells A, et al. A Mobile Health Solution Complementing Psychopharmacology-Supported Smoking Cessation: Randomized Controlled Trial. JMIR mHealth and uHealth. 2020 Apr 27;8(4):e17530.
16. 16)Masaki K, Tateno H, Nomura A, Muto T, Suzuki S, Satake K, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. npj Digital Medicine [Internet]. 2020 Mar 12;3(1):1–7.
17. 17))Faro JM, Chen J, Flahive J, Nagawa CS, Orvek EA, Houston TK, Allison JJ, Person SD, Smith BM, Blok AC, Sadasivam RS. Effect of a Machine Learning Recommender System and Viral Peer Marketing Intervention on Smoking Cessation: A Randomized Clinical Trial. JAMA Netw Open. 2023 Jan 3;6(1):e2250665. doi: 10.1001/jamanetworkopen.2022.50665. PMID: 36633844; PMCID: PMC9856644
18. 18)Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health. 2019 Sep 30;5:2055207619880676. doi: 10.1177/2055207619880676. Erratum in: Digit Health. 2020 May 22;6:2055207620930958. doi: 10.1177/2055207620930958. PMID: 31620306; PMCID: PMC6775545.
19. 19)Leeuwis L, He L. Hi, I'm Cecil(y) the Smoking Cessation Chatbot: The Effectiveness of Motivational Interviewing and Confrontational Counseling Chatbots and the Moderating Role of the Need for Autonomy and Self-Efficacy. In Følstad A, editor, Chatbot Research and Design. CONVERSATIONS 2022. Lecture Notes in Computer Science. Vol. 13815. Springer Cham. 2023. p. 3-17 doi: 10.1007/978-3-031-25581-6_1

20. 20)Heffner JL, Kelly MM, Reilly ED, Reece SG, Claudio T, Serfozo E, Baker K, Watson NL, Karekla M. An Avatar-Led Web-Based and SMS Text Message Smoking Cessation Program for Socioeconomically Disadvantaged Veterans: Pilot Randomized Controlled Trial. JMIR Form Res. 2023 Apr 14;7:e44503. doi: 10.2196/44503. PMID: 37058346; PMCID: PMC10148208.
21. 21)Karekla M, Savvides SN, Gloster A. An Avatar-Led Intervention Promotes Smoking Cessation in Young Adults: A Pilot Randomized Clinical Trial. Ann Behav Med. 2020 Oct 1;54(10):747-760. doi: 10.1093/abm/kaaa013. PMID: 32383736.
22. 22)Almusharraf F, Rose J, Selby P. Engaging Unmotivated Smokers to Move Toward Quitting: Design of Motivational Interviewing–Based Chatbot Through Iterative Interactions. Journal of Medical Internet Research. 2020 Nov 3;22(11):e20251.
23. 23)Bricker JB, Sullivan B, Mull K, Santiago-Torres M, LavistaFerres JM. Conversational Chatbot for Cigarette Smoking Cessation: Results From the 11-Step User-Centered Design Development Process and Randomized Controlled Trial. JMIR MhealthUhealth. 2024 Jul 23;12:e57318. doi: 10.2196/57318. PMID: 38913882; PMCID: PMC11303891.
24. 24)Saiyed A, Layton J, Borsari B, Cheng J, Kanzaveli T, Tsvetovat M, Satterfield JM. . Technology-Assisted Motivational Interviewing: Developing a Scalable Framework for Promoting Engagement with Tobacco Cessation Using NLP and Machine Learning. Procedia Computer Science. 2022; (206):121-131.
25. 25)Brown A, Kumar AT, Melamed O, Ahmed I, Wang YH, Deza A, Morcos M, Zhu L, Maslej M, Minian N, Sujaya V, Wolff J, Doggett O, Iantorno M, Ratto M, Selby P, Rose J. A Motivational Interviewing Chatbot With Generative Reflections for Increasing Readiness to Quit Smoking: Iterative Development Study. JMIR Ment Health. 2023 Oct 17;10:e49132. doi: 10.2196/49132. PMID: 37847539; PMCID: PMC10618902.
26. 26)Lakatos R, Pollner P, Hajdu A, Joó T. A multimodal deep learning architecture for smoking detection with a small data approach. Front ArtifIntell. 2024 Feb 28;7:1326050. doi: 10.3389/frai.2024.1326050. PMID: 38481821; PMCID: PMC10936563.
27. 27)He L, Basar E, Krahmer E, Wiers R, Marjolijn Antheunis. Effectiveness and User Experience of a Smoking Cessation Chatbot: Mixed Methods Study Comparing Motivational Interviewing and Confrontational Counseling. Journal of Medical Internet Research. 2024 Aug 6;26:e53134–4.
28. 28)He L, Basar E, Wiers RW, Antheunis ML, Krahmer E. Can chatbots help to motivate smoking cessation? A study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. BMC Public Health. 2022 Apr 12;22(1):726. doi: 10.1186/s12889-022-13115-x. PMID: 35413887; PMCID: PMC9003955.
29. 29)Loveys K, Lloyd E, Sagar M, Broadbent E. Development of a Virtual Human for Supporting Tobacco Cessation During the COVID-19 Pandemic. J Med Internet Res. 2023 Dec 5;25:e42310. doi: 10.2196/42310. PMID: 38051571; PMCID: PMC10731553.
30. 30)Singh T, Truong M, Roberts K, Myneni S. Sequencing conversational turns in peer interactions: An integrated approach for evidence-based conversational agent for just-in-time nicotine cravings intervention. Digit Health. 2024 Feb 13;10:20552076241228430. doi: 10.1177/20552076241228430. PMID: 38357587; PMCID: PMC10865956.
31. 31)Küçükali H, Erdoğan MS. AI for Tobacco Control: Identifying Tobacco-Promoting Social Media Content Using Large Language Models. Nicotine Tob Res. 2025 May 22;27(6):988-996. doi: 10.1093/ntr/ntae276. PMID: 39579342; PMCID: PMC12095795.
32. 32))Bendotti H, Lawler S, Ireland D, Gartner C, Marshall HM. Co-Designing a Smoking Cessation Chatbot: Focus Group Study of End Users and Smoking Cessation Professionals. JMIR Hum Factors. 2024 Aug 19;11:e56505. doi: 10.2196/56505. PMID: 39159451; PMCID: PMC11369547
33. 33)Hernández-Llanes NF, Sánchez-Domínguez R, Álvarez-Reza S, Fernández-Cáceres C, Marín-Navarrete R. Factors Associated with Tobacco Cessation Services Request Among Users of an Online Self-Screening Questionnaire. Subst Use Misuse. 2025;60(4):604-610. doi: 10.1080/10826084.2024.2445851. Epub 2024 Dec 28. PMID: 39731740.
34. 34)P. Roy, F. Hossain and N. Jahan, "Machine Learning Approach to Predict Influence of Smoking on Student Life," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT); 2021 Jul 6-8; Kharagpur, India. IEEE; 2021. P.1-6.
35. 35)Engelhard MM, Oliver JA, Henao R, Hallyburton M, Carin LE, Conklin C, McClernon FJ. Identifying Smoking Environments From Images of Daily Life With Deep Learning. JAMA Netw Open. 2019 Aug 2;2(8):e197939. doi: 10.1001/jamanetworkopen.2019.7939. PMID: 31373647; PMCID: PMC6681554.
36. 36)Issabakhsh M, Sánchez-Romero LM, Le TTT, Liber AC, Tan J, Li Y, Meza R, Mendez D, Levy DT. Machine learning application for predicting smoking cessation among US adults: An analysis of waves 1-3 of the PATH study. PLoS One. 2023 Jun 8;18(6):e0286883. doi: 10.1371/journal.pone.0286883. PMID: 37289765; PMCID: PMC10249849.
37. 37)Kato A, Tanigawa T, Satake K, Nomura A. Efficacy of the Ascure Smoking Cessation Program: Retrospective Study. JMIR MhealthUhealth. 2020 May 14;8(5):e17270. doi: 10.2196/17270. PMID: 32406856; PMCID: PMC7256743.
38. 38)Lai CC, Huang WH, Chang BC, Hwang LC. Development of Machine Learning Models for Prediction of Smoking Cessation Outcome. Int J Environ Res Public Health. 2021 Mar 5;18(5):2584. doi: 10.3390/ijerph18052584. PMID: 33807561; PMCID: PMC7967540.
39. 39)Wang X, Zhao K, Cha S, Amato MS, Cohn AM, Pearson JL, Papandonatos GD, Graham AL. Mining User-Generated Content in an Online Smoking Cessation Community to Identify Smoking Status: A Machine Learning Approach. Decis Support Syst. 2019 Jan;116:26-34. doi: 10.1016/j.dss.2018.10.005. Epub 2018 Oct 15. PMID: 31885411; PMCID: PMC6934371.
40. 40) Kim N, Loh WY, McCarthy DE. Machine learning models of tobacco susceptibility and current use among adolescents from 97 countries in the Global Youth Tobacco Survey, 2013-2017. PLOS Glob Public Health. 2021 Dec 8;1(12):e0000060. doi: 10.1371/journal.pgph.0000060. PMID: 36962106; PMCID: PMC10021689.
41. 41)Flanagan SV, Vargas A, Smith J. Application of machine learning techniques to profile smoking behavior of adolescent girls in Ghana. Gates Open Res. 2024 Nov 20;8:2. doi: 10.12688/gatesopenres.14991.2. PMID: 40881194; PMCID: PMC12381352.
42. 42)Siegel LN, Wiseman KP, Budenz A, Prutzman Y. Identifying Patterns of Smoking Cessation App Feature Use That Predict Successful Quitting: Secondary Analysis of Experimental Data Leveraging Machine Learning. JMIR AI. 2024 May 22;3:e51756. doi: 10.2196/51756. PMID: 38875564; PMCID: PMC11153975.
43. 43)Hébert ET, Suchting R, Ra CK, Alexander AC, Kendzor DE, Vidrine DJ, Businelle MS. Predicting the first smoking lapse during a quit attempt: A machine learning approach. Drug Alcohol Depend. 2021 Jan 1;218:108340. doi: 10.1016/j.drugalcdep.2020.108340. Epub 2020 Oct 11. PMID: 33092911; PMCID: PMC8496911.
44. 44)Suchting R, Hébert ET, Ma P, Kendzor DE, Businelle MS. Using Elastic Net Penalized Cox Proportional Hazards Regression to Identify Predictors of Imminent Smoking Lapse. Nicotine Tob Res. 2019 Jan 4;21(2):173-179. doi: 10.1093/ntr/ntx201. PMID: 29059349; PMCID: PMC7962780
45. 45)Caccamisi A, Jørgensen L, Dalianis H, Rosenlund M. Natural language processing and machine learning to enable automatic extraction and classification of patients' smoking status from electronic medical records. Ups J Med Sci. 2020 Nov;125(4):316-324. doi: 10.1080/03009734.2020.1792010.

Epub 2020 Jul 22. PMID: 32696698; PMCID: PMC75948654
46. 46)Rijhwani K, Mohanty VR, Yb A, Singh V, Hashmi S. Applicability of Data Mining and Predictive Analysis for Tobacco Cessation: An Exploratory Study. Front Dent. 2020 Oct 20;17:24. doi: 10.18502/fid.v17i24.4624. PMID: 36042799; PMCID: PMC9375116.
47. 47)Alphonse A, Stewart K, Brown J, Perski O. Exploring Users' Experiences With a Quick-Response Chatbot Within a Popular Smoking Cessation Smartphone App: Semistructured Interview Study. JMIR Form Res. 2022 Jul 7;6(7):e36869. doi: 10.2196/36869. PMID: 35797093; PMCID: PMC9305398.
48. 48))Davagdorj K, Lee JS, Pham VH, Ryu KH. A Comparative Analysis of Machine Learning Methods for Class Imbalance in a Smoking Cessation Intervention. Applied Sciences. 2020 May 9;10(9):3307.
49. 49)Harris DR, Henderson DW, Corbeau A. Improving the Utility of Tobacco-Related Problem List Entries Using Natural Language Processing. AMIA Annu Symp Proc. 2021 Jan 25;2020:534-543. PMID: 33936427; PMCID: PMC8075422.
50. 50)Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile Applications for the Treatment of Tobacco Use and Dependence. Curr Addict Rep. 2019 Jun;6(2):86-97. doi: 10.1007/s40429-019-00248-0. Epub 2019 May 9. PMID: 32010548; PMCID: PMC6994183
51. 51)Whittaker R, Dobson R, Garner K. Chatbots for Smoking Cessation: Scoping Review. J Med Internet Res. 2022 Sep 26;24(9):e35556. doi: 10.2196/35556. PMID: 36095295; PMCID: PMC9514452.
52. 52)He L, Balaji D, Wiers RW, Antheunis ML, Krahmer E. Effectiveness and Acceptability of Conversational Agents for Smoking Cessation: A Systematic Review and Meta-analysis. Nicotine Tob Res. 2023 Jun 9;25(7):1241-1250. doi: 10.1093/ntr/ntac281. PMID: 36507916; PMCID: PMC10256885.
53. 53)Bendotti H, Lawler S, Chan GCK, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: A systematic review and meta-analysis. Digit Health. 2023 Nov 3;9:20552076231211634. doi: 10.1177/20552076231211634. PMID: 37928336; PMCID: PMC10623979.
54. 54)Bickel WK, Tomlinson DC, Craft WH, Ma M, Dwyer CL, Yeh YH, Tegge AN, Freitas-Lemos R, Athamneh LN. Predictors of smoking cessation outcomes identified by machine learning: A systematic review. Addict Neurosci. 2023 Jun;6:100068. doi: 10.1016/j.addicn.2023.100068. Epub 2023 Jan 31. PMID: 37214256; PMCID: PMC10194042..

Appendix-1 Detailed search strategy

1. Pubmed

No.	Query	Results
#3	# 1 AND #2	1552
#2	(((((("Artificial Intelligence"[Mesh] OR "Generative Artificial Intelligence"[Mesh]) OR ("Machine Learning"[Mesh] OR "Machine Learning Algorithms"[Mesh])) OR ("Transfer Machine Learning"[Mesh] OR "Representation Machine Learning"[Mesh] OR "Unsupervised Machine Learning"[Mesh] OR "Supervised Machine Learning"[Mesh] OR "Boosting Machine Learning Algorithms"[Mesh] OR "Reinforcement Machine Learning"[Mesh] OR "Extreme Learning Machines"[Mesh] OR "Federated Learning"[Mesh])) OR ("Deep Learning"[Mesh] OR "Detection Algorithms"[Mesh])) OR ("Natural Language Processing"[Mesh] OR "Large Language Models"[Mesh])) OR ("Neural Networks, Computer"[Mesh] OR "Convolutional Neural Networks"[Mesh] OR "Radial Basis Function Networks"[Mesh] OR "Recurrent Neural Networks"[Mesh] OR "Feedforward Neural Networks"[Mesh] OR "Graph Neural Networks"[Mesh])) OR "Support Vector Machine"[Mesh]) OR "Random Forest"[Mesh]) OR "Large Language Models"[Mesh] OR "Avatar"[Majr] OR "Predictive Learning Models"[Mesh] OR chat[tiab] OR chatbot* OR "chat bot*" OR "artificial intelligence" OR ai[ti] OR "conversation* agent*"[tiab] OR "conversation* dialog*"[tiab]	31,853
#1	(((("Nicotine"[Mesh]) OR "Tobacco Use Disorder"[Mesh]) OR ("Smoking"[Mesh] OR "Smoking Cessation"[Mesh] OR "Smoking Prevention"[Mesh] OR "Smoking Reduction"[Mesh] OR "Tobacco Smoking"[Mesh] OR "Cigarette Smoking"[Mesh] OR "Cigar Smoking"[Mesh])) OR	7505

	"Tobacco Control"[Mesh]) OR "Tobacco Products"[Mesh] OR "Tobacco, Smokeless"[Mesh]) OR "Catha"[Mesh]) OR "Tobacco Use"[Mesh]) OR "Tobacco, Waterpipe"[Mesh]	

2.Google Scholar

Search Term	**Search Yield**
Artificial intelligence in tobacco| Machine learning in tobacco| Deep learning in tobacco| Chatbot in tobacco	6080

Due to Google Scholar's interface & relevance -ranking algorithm, we screened only the first 200 results as recommended by Bramer et al,2018 in their published guidance for systematic reviews. The screening was conducted in 2 stages: 1)tiltle screening-to exclude completely irrelevant studies 2)Abstract screening to recognize studies that fulfill our inclusion criteria. The remaining studies were exported to Rayyan.

3.The Cochrane Review- 6 studies

4. DOAJ- 30 studies

Appendix 2-Summary of included studies evaluating applications of AI methods in predicting tobacco use.

				AI Intervention	**Focus of the Study**		
Pranta Roy 2021	analytical Cross-sectional study	500	Undergraduates(91.9%), Postgraduates (8.1%)	Supervised ML algorithms (KNN, Random Forest & SVM)	Smoking status & its impact on academic life	Students smoking data(from google forms&questionnaire) was used to train(75%)the model and then testit(25%).	Most accurate- KNN & RF, Most precise- SVM, Best Recall- RF.
Xi Wang 2019	Secondary data analysis of Retrospective Cohort study	2120 community post	NA	Supervised ML(NB, Logistic regression, J48 decision tree, SVM, AdaBoost)	Smoking status	Community post were classified into- +class(not smoking) and -class(smoking and unclear status) &then 5 ML models evaluated them for unigrams, special phrases, replies &thread	Model5 with AdaBoost using J-48 decision tree

						details etc	
Nayoung Kim 2021	Secondary data analysis (of a cross-sectional survey)	343, 481 adolescents	Age13-15 yrs, most were exposed to second hand smoke(SHS) and tobacco promotion	Supervised ML (GUIDE analysis)	Susceptibility and current use of tobacco	Adolescent data from GYTS(2013&17) was cluster sampled & divided into:tobacco naive, non susceptible/ susceptible &current tobacco user	important predictors of susceptibility & current use to tobacco use as shown byGUIDE-exposure to tobacco industry promotions , SHS exposure,support for outdoor&indoor SFPs etc
Sara V. Flanagan 2024	Secondary data analysis (of stepped wedge cluster RCT)	9000	adolescent girls aged13-19 yrs from Ghana	Supervised ML(RF classifier)	Smoking status	Data from girls was used by basic RF classifier to identify their smoking status which was trained using ROS,ADASYN,SMOTE etc	Best performing model:SMOTE with RF classifier
Robert Lakatos 2024	developmental +experimental(lab based)	5 short advertising videos	videos<2minat 240 resolution; contained average of 13 sec of smoking	multimodal deep learning(CLIP-ViT-B-32 &EfficientNet B5)	to detect smoking-related content in text and images.	Selected mp4 videos, break them into frames then filtered using CLIP-ViT, converted images into vectors and compared them to smoking related vector. Also used Efficient Net B5 version for filtering.	Most accurate-combined(detected 4 sec of smoking)> CLIP-ViT-B-32(25 sec)>EfficientNet B5(28 sec). For text-best was XLM-RoBERTa
Andrea Caccamisi	Secondary data analysis with a	337, 409	EMR data 2013,2014	Supervised ML using NLP	Smoking	EMR data categorizin	Best predicting

2020	methodological focus	rows of smoking related text in EMRs	studies of Sweden with smoking status categories		status	g smoking status as - current, ex and non smoker or unknown was used to train & test ML-NLP models like SMO,KNN ,NB,J-48	model- SMO using unigrams & bigrams to understand the sentences
Cheng-Chien Lai 2021	Retrospective Cohort study	n=4875	mean age=46.7 yrs, Avg CPD=20.1	Supervised ML(SVN,RF,LR, KNN ,CART,NB)+deep learning(ANN)	Success rate of smoking cessation	7ML algorithms were trained on Python3.7 using 4375 participants dataset and then tested on dataset of 500 for prediction of abstinence.	Best predictor model- ANN
Mona Issabakhsh 2023	Prospective Cohort Study	9281	Current established smokers from PATH survey who are >18 yrs,smoked >100 cig in lifetime	Supervised ML(RF,GBM,GLM,XGBoost)	Success rate of smoking cessation	Training & testing of ML algorithms to correctly predict quit status was done using 181 PATH survey variables	Best models- GBM and RF .Top variablesselected : age at smoking initiation, past 30 days polytobacco product use etc
Lara N 2020	secondary data analysis (of existing cohort study)	161	>18yrs, >=16 cig/ day, meetscriteria for nicotine dependence	Supervised ML (CART analysis)	Success rate of smoking cessation	Six,60 min CBT sessions to 2 cohort of smokers(Training cohort:8 weeks of nicotine patch and validation cohort-no NRT) & then ML decision tress was used to predict outcome; cessation measured by	delay discounting (DDF) significantly predicts smoking cessationoutcomes better than chance, in both training and validation samples.

						CO & urine COT levels	
Leeann Nicole Siegel 2024	secondary analysis of RCT data	N=133	Mean age=45.6 yrs ;mean score on Fragerstrom test=4.8	supervised machine learning (logistic regression)	Identify patterns of Smoking Cessation App Feature Use for quitting	Subjects of both groups(incentivized-62 &non-incentivized-71) used quit start app & then SML was used to predict cessation from 28 variables of participants; Cessation by self reported surveys	3 app features identified-1) I slipped button 2) users who viewed 'tips'. FYI's etc more and completed more badges(more likely)
Kavita Rijhwani 2020	Secondary Data Analysis(ofRetrospective Cohort Study)	655	users who visited tobacco cessation clinic between 2015 to 2016	supervised ML (NB, SMO, RF, J48, decision stump)	To detect quit status and its predictors	Patients' data was uploaded to diff algorithms of ML for prediction of quit status	Decision stumps&SMO- most correctly classified instances while RF & NB most incorrectly. Predictors ranking:previous attempt> type of intervention etc
Norberto Francisco 2024	descriptive cross-sectionalstudy	14182	descriptive cross-sectional	Supervised machine learning (random forest algorithm)	Identifying factors for requesting tobacco cessation services	Users filed tobacco questionnaire, leaving their contact info & received address of smoking cessation t/t centre	RF correctly predicted 78.6% users who requested cessation services. Top predictors-older age, dependence severity etc
KhishigsurenDavagdorj 2020	Retrospective observational study	3692	former smokers= 951 and current smokers =2741	supervised machine learning(LR,RF,GBT,KNN,SVM)+Deep learning(MLP	Handling class imbalance in prediction of	Data of smokers from Korean National Health Survey taken. First Lasso used(excluded irrelevant	Best predictor overall-SMOTE with GBT classifier. In male SMOTE with RF

					smoking cessation	features), multi-collinearity analysis, SMOTE &ADASYN(to correct oversampling)then compared ML classifiers topredict successful quitting	classifier. In female ADASYN with MLP
Robert Suchting 2019	Secondary data analysis (of RCT with active control)	92 smokers	Mean age=51.9 yrs, mean cig/day= 18,oncessation program safety-net hospital Texas.	Supervised ML(Elastic Net Cox Regression, proportional hazards regression)	moment-to-moment predictors of smoking lapse	Smokers randomized to-usual care group(pharma+group counselling)contingency management group(usual care+small financial incentives). 26 candidate predictors from EMA via smartphones used to predict lapse	Elastic Net cox regression model retained 17 out of 26 predictors Approximated model retained 6 predictors-feeling irritable,having cigavailable etc
Emily T. Hébert 2021	Secondary analysis of three arm pilot RCT	74	>18 yrs age,avg smoking=21.9 cig/day, willing to quit7 days from their first visit.	Supervised ML (JITAI+Elastic Net Regression)	Identifying factors for predicting lapse	3 groups-Smart T2(for smoking lapse risk) Quitguideapp(to track cravings, triggers) & usual care. EMA data was then utilized by ML algorithm to predict persons' first lapse	Top3- self reported odds, motivation & confidence to avoid smoking

Appendix 3 -Summary of included studies evaluating applications of AI methods in tobacco abstinence/craving/promotion

Study Author	Study design	Sample size	Participant characteristics	AI Intervention	Control	Study Strategy	Outcome
Eduardo Olano Espinosa 2022	Pragmatic, Multicentre RCT	513	>18 yrs, smoked at least 1cig in last month	Hybrid AI model (Rule based NLP chatbot + Bayesian probabilistic decision process	yes	CG (n=271, LTFU=92) - received behavioral therapy+ pharmaIG(n=242,LTFU=87) same via chatbot	CAR at 6 months, QOL, mean interaction time, no of controls &intensive use were more improved in IG; abstinence self reportedverified by CO oximetry<10 ppm
Fahad Almusharraf 2020	iterative developmental design	121	>18 yrs, 10.1% of patientsshowedmoderate or high dependence (HSI scores)	Hybrid AI(rule based NLP chatbot system+supervised ML)	NA	Participant conversed with chatbot & gave feedback.6568 responses divided into 2 categories- smoking good or bad(original 10 then additional 11)	Recall &F1 score of SR2(developed from Grp1 using spacy-NLP from grp 11) was better than SR1(used in original 10)
Linwei He 2024	Mixed-Methods study: RCT + Descriptive, Cross-sectional survey of semi-structured interviews	287	mean age=21.1 yrs;Mean Fagerström Test score = 1.17(overall low dependence)	rule based NLP chatbot	yes	2 groups created: MI style(n=145,LTFU =33) ; CC style(n=142,LTFU =25) and 16 subjects were randomly selected for final interview	Overall user experience(interaction satisfaction, engagement, empathy) was better for MI style. Intention to quit &self efficacy had no diff.
Laura Carassco 2020	parallel,single blind(participants), active control RCT	240	>18 yr,average daily cigin IG and CG was 21.45 and 20.75	AI driven smartphone app(HRS)	yes	IG(n=120,LTFU=69) received So-Lo-M0 app with pharmacotherapy ;CG(n=120;LTFU=75)-only pharma	HRS group showed greater efficacy for achieving 1 year tobacco abstinence(measured by exhaled CO &urine cotinine)
Katsunori Masaki 2020	multicentre, parallel single blind(participants), with active control RCT	572	mean age=46 yrsMedian of pack years=20	Rule based CA	yes	CASC grp(n=285,LTFU at 52 weeks=40) included digital entries, chatbot, educational videos &CO checker : Control grp(n=287 LTFU at 52 weeks=42) received appointment	CASC system significantly improved long term CAR(biochemically validated by mobile CO checker), time to first lapse after

						reminders	quit date, MSS score, withdrawal symptoms (FTCQ-12) &cravings(KTSND)
Alice Alphonse 2022	qualitative research study	14	age >18 yrs. 2groups- a)past year smokers and had used Quit coach (n=5) b)current smokers who agreed to used Quit coach(n=9)	Rule based chatbot	No	Users were provided with online one-to-one semi-structured, qualitative interviews with the chatbot regarding smoking experience	QuitCoach's anthropomorphism, accountability, interaction style and check-ins during users need increased engagement with chatbot
Jamie M. Faro 2023	Multi arm parallel RCT with partial allocation,single blind(staff)	1487	> 18 yrs, with moderate to heavy daily cigarette use.	supervised ML (ML recommender system)	yes	Group A-ML recommender and viral tool kit(n=371) GroupB- only ML recommender(n=370) Group C- std messaging and viral tool kit(n=374) GroupD- only std messaging(n=372). The std messaging system delivered messagesbased on only baseline motivation to quit whereas ML used feedback from users(LTFU=46.1%)	Quit rate at 6 month- 1)Group A >Group B(p=.005) 2)those exposed to viral tool kit- Group A and C >not exposed(p=.01). No baseline difference in ML recommender- Group A and B vs std messaging -Group C and D(p=0.16); Abstinenceself reported validated by saliva cotinine kits
Olga Perski 2019	Parallel, Open-label RCT with unequal randomization & active control	57,214 smokers	daily or non daily smokers aged> 18 yrs whoset a quit date <2 days before &<14 days after registration	rule based CA	yes	Smokers randomized into 4:1- CG(n=51875LTFU=90.3%), had pro version of smoke free app, Intervention group (n=5339,LTFU=80.1%) had pro version of smoke free app+ supportive , AI driven chatbot.	Engagement(median logins), follow up rate, and quit success at 1 month was significantly more in IG, Abstinence-self reported
Lotte Leeuwis 2023	parallel, open label, two arm RCT	233	>18 yrs, smokedatleast 1 cigarette in the week prior to participation	rule based chatbot	yes	Group 1-MI style chatbot Cecily(n=121) Grp2- CC style chatbot Cecil (n=112) LTFU 37	No significant effect of chatbot communication styleon intention to quitand user satisfaction.

Jonathan B Bricker 2024	Iterative Developmental +pilot RCT	n=404 +n=14	age>18 yrs, having smoked atleast 1 cigarette/day for past 12 months and wanting to quit in next 14 days	Hybrid AI model(rule based NLP chatbot+LLM)	yes	Quitbot developed and tested on 404 participants who were randomised into-Quitbot(n=200), Smokefree TXT(n=149) and placebo/delay access control group(n=55). After users feedback, in 11th step free form Q&A were added.	Engagement &30 dayPPA-Quitbot users>SFT >placebo.For Final version of Quitbot-participants were highly receptive for stuctured and free form chats. Abstinence-self reported
Ayaka Kato 2020	Retrospective cohort	177 adult smokers	mean age=44.6 yrs, mean cigarettes/day-16.2	rule based chatbot	no	Participants enrolled in 24 weeks programgiven Ascure app comprised of educational videos, exercises , digital diary, chatbot sessions etc	Predictor of CAR at week 21-24 -no of days digital diaries written and no ofvideo tutorials watched in first 12 weeks; Abstinence-self reported validated by salivary cotinine test
Linwei He 2022	Mixed-Methods study: RCT + Descriptive, Cross-sectional survey of semi-structured interviews	153 smokers	average daily cigarette consumption & average years of smoking be 5.1 & 4.2 respectively	Rule based CA	yes	2 sessions of assessment were held with both styles of chatbot-MI(n=78) & neutral style-(n=75)	No sig difference wqas seen for engagement, therapeutic alliance & perceived empathy for both styles. Motivation to quit smoking & perceived communication competence showed positive results, irrespective of chatbot style.
Jaimee L Heffner 2023	Pilot parallel double blinded RCT with an active control	49	mean age=51.3 yrs, smoked average of 16.7	Embodied CA	yes	2 groups-VetFlexiquit delivered ACT(n=25; LTFU=8) & SmokeFree Vet delivered educational materials(n=24 LTFU=5)	At 3 months-Satisfaction & Utilization(average log-ins)for VetFlexiquit users more; Cessation(self reported verified by saliva cotinine)-no

							significant diff
Ahson Saiyed 2022	Iterative Developmental Study+Pilot clinical trial	34	14016 utterances from 81 subreddits, 7702 hand-annotated utterances by volunteers, 34 smokers (to test TAMI)	Hybrid AI model: rule based chatbot+ deep learning based NLP models (BERT, RoBERTa, XLNet, GCN, GAN)	NO	Developed & tested TAMI using RASA framework which had 3 modules (onboarding, MI for those not ready to quit % outboarding for those ready to quit)	TAMI successfullycompleted tasks of smoking cessation screening, counselling& referral. Best model to train TAMI for utterance classification - RoBERTa,& topic classification is XLNet.
Maria Karekla 2019	parallel,single blinded(participants)RCT	84	mean age=22.44yrs,smoking >_1cigarette/day during past 30 days	Embodied CA	yes	Grp1- avatar led ACT via Flexiquit(n=49;LTFU=21) 2Grp 2 to control group(n=35;LTFU =7) Cessation at 7 day point prevalence and no of cigarette smoked/day assessed	Flexiquit showedstatistical significant increase in 7 day PPA, self efficacy& intention-to-quit & decrease in avg no. of cigarette smoked/day & nicotine dependence.
Andrew Brown 2023	iterative developmental	349	mean age=30 yrs, mean =10.8 cig /day	Hybrid AI model(chatbot+ NLP+ deep learning ML)	no	4 versions of chatbot made- MIBot v4.7 asked questionsMIBot v5.0- asked ques and used initial version of reflectionMIBot v5.1-used improved reflection generator MIBot v5.2- used extended interactions.	All 4 MIBot versions statistically improved confidence to quit. MIBotv5.2 came out to be superior in terms of perceived empathy & readiness to quit
Kate Loveys 2023	mixed method descriptive cross sectional +qualitative thematic analysis	115	Users from 49 counties, englishspeaking , mostly from India	Embodied Conversational Agent	No	Florence delivered cessation content & helped users to make quit plan for which the users provided feedback.	Majority users reported good feedback: good info, clear communication & connection.That could be improved

							were- more personalisation and interactivity and more human like voice.
Tavleen Singh 2024	Mixed-Methods study	1000 subset of messages	NA	Hybrid AI model (Rule based chatbot with supervised ML & deep learning models)	No	Extracted relevant nicotine cravings related scenariosfromQuit Net real peer conversations &mapped them in architecture of CA using RASA's NLU& DM	Created 12 user intents (5 based on extracted scenarios & 7 to support conversations) for training of NLU & DM models to develop a CA for just-in-time cravings.
Hollie Bendotti 2024	Mixed-Methods study (descriptive, cross-sectional with beta-testing process)	9 + 7	Smoking cessation professionals- most were previous smokers, End-userscurrently smoked, 2 had quit	Rule based Chatbot with limited NLP capabilities	No	Participants tested the Quin app 1-2 weeks prior & then participated in Focus group discussion which was analysed(positive /negative/suggestions) and feedback was incorporated in Quin.	4 majorfeedback were- inability to act in "Moment of crisis", lack of customisation, limitedhuman like characteristics, positive feedback as cessation tool. These were used to update Quin (from version 21 to 28)
Hüseyin Küçükali 2024	Mixed-method study(qualitative then quantitative)	N=177, 684 Tweets	avg=6253.2 tweets/day, 37.6% were original tweets, 37.8% retweets and 24.7% were replies	Deep learning models (LLM)	No	5000 tweets (for/against tobacco or both) divided into training and test dataset(4:1) which was coded by BERT to create 2 models-Model 1(all tweets) & Model 2 (consistent review)	Model 2 performed better than Model 1in terms of accuracy, precision & overall performance in describing whether social media content promotes tobacco or not.
Matthew M. Engelhard 2019	Analytical, Cross-sectional study	4902 images by 169 smokers	>18 years,active smokers (≥5 cigarettes/day for ≥1 year	ML models using deep learning (CNN) & supervised ML models	No	Images of 4smoking and 4 non smoking locations from each smoker were classified using hybrid AI i.eInception 4CNN	Model predictions were strongly correlated with self reported craving.Obje

				(logistic regression)		(to detect object in images)and LR (to classify images)	cts associated with smoking env. weremoving van, park bench, gas pump, park bench and those with non smoking environment were library, grocery store, church etc
Daniel R. Harris 2020	Retrospective, observational, methodological study	1667 enteries	1667 problem list tobaccoenteries(2017-19) from University of Kentucky's OPD HER	Hybrid AI (supervised ML based CLAMP-NLP + rule based AI like NegEx)	No	These entries were classified by CLAMP-NLP: Tobacco (n=1499), Former tobacco (n=102) &Non-tobacco (n=66)	Achieved precision up to 0.99for tobacco users, 1.0 for former, 0.96 for non-users after post-processing. Also createdsimple rules (3 for non-tobacco problems, 12 for former-tobacco & 10 for current tobacco use) for adjusting class assigned to each problem

Appendix 4: Summary of review-level studies

Study Author	Study Design	Inclusion Criteria	Outcome
Roger Vilardaga 2019	Narrative review	1) reported research on a smoking cessation app 2)peer reviewed 3)indexed in PubMed or google scholar	33 apps-1)almost 1/2(n=16) targeted smoking in general population{eg.Clickotine,SSC app, geolocation apps,CBT-Quit Genius etc2)remaining half(n=17) targeted smoking in specific population: female smokers(SMSF, Smoke free baby)etc. Majority of apps did not use Machine Learning and other complex AI tools
Linwei He MSc 2024	Systematic Review and meta-analysis	1)peer reviewed 2)reported use of CA or synonymous system 3)address smoking cessation or relapse prevention 4)reported outcomes from direct end users	Out of 13 studies- 2 studies-low risk of bias, 5-moderate &5 had a high risk of bias(due to unequal amount of intervention received &self reported abstinence) Meta-analysis of 6RCT showedCA significantly increased the odds of abstinence comparedto control groups(p<0.001).Examinationof the funnel plot indicatedpossible publication bias. The remaining non-RCTs with pre-post designs also

			reported positive results.
Hollie Bendotti 2023	Systematic review and meta-analysis	1)RCT that involved conversational AI cessation intervention compared with control group including no intervention, usual care and active comparator. 2)Bidirectional communication(AM was excluded)	5RCTS-Meta-analysis of abstinence outcomes(n=3) at 6-month f/u (n=3) found that AI intervention were significantlymore likely to quit smoking compared tocontrol(p< 0.001). The test of funnel plot symmetry didnot provide any evidence of asymmetry. Overall, the risk of bias was high across the studies
Robyn Whittaker 2022	Scoping Review	1) conducted in tobacco smokers, 2) conducted between 2000 & 2021, 3) available in English, 4) included a chatbot intervention	10 studies-Most studies had positive result regarding CA but most studies wereof low quality. Only 2 RCTs were found. In 1 RCT, only 10.7% of the overall sample responded at a 1-month follow-up. The other RCT failed to report any P values. In addition, the follow-up period was short (1 to 2 months).
Warren K. Bickel 2023	Systematic Review	1) Clinical trials (randomised or not) & observational studies evaluating smoking cessation outcomes using ML, 2) published between 2006 & 2022	12 studies-Allemployed random forests, classification trees, regression trees or logistic regression.. 3) Best predictors identified by ML: delay discounting.

Appendix 5: List of abbreviations:
AI- Artificial Intelligence
ANN- Artificial Neural Network
ADASYN- Adaptive Synthetic Sampling
BERT- Bidirectional Encoder Representations from Transformers
BMI- Body Mass Index
CA- Conversational Agent
CART- Classification and Regression Trees
CASC- Cureapp Smoking Cessation
CNN- Convolutional Neural Network
CLAMP- Clinical language annotation, modeling and processing
DOAJ- Directory of Open Access Journals
DL- Deep Learning
ENR- Elastic Net Cox Regression
EHR- Electronic Health Record
EMR- Electronic Medical Record
GAN- Generative Adversarial Network
GATS- Global Adult Tobacco Survey
GBM-Gradient Boosting Machine
GLM- Generalised Linear Model
GBM-Gradient Boosting Machine
GBT- Gradient Boosting Trees
GLM- Generalised Linear Model
GUIDE- Generalised, Unbiased Interaction Detection and Estimation
GYTS- Global Youth Tobacco Survey
JITAI- Just- In-Time Adaptive Intervention
KNN- K-Nearest Neighbors
LLM- Large Language Model

LSTM- Long short-term memory
LR- Linear Regression
LMICs- Low and Middle Income countries
LTFU- Loss to follow up
ML- Machine Learning
MLP- Multilayer Perceptron
NLP- Natural Language Processing
NLU- Natural Language Understanding
NB- Naïve Bayes
NRT- Nicotine Replacement Therapy
PRISMA-ScR- Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews
PCC- Population, Concept, Context
PATH Survey- Population Assessment of Tobacco and Health Survey
RCT- Randomized Control Trial
RF- Random Forest
RNN- Recurrent Neural Network
SLM- Small Language Model
SML- Supervised Machine Learning
SVM- Support Vector Machine
SMOTE-Synthetic Minority Oversampling Technique
SFP- Smoke free policy
SMO- Sequential Minimal Optimization
TC- Training Cohort
TAMI- Technology Assisted Motivational Interviewing coach
VC- Validation Cohort
WEKA- Waikato Environment for Knowledge Analysis
WHO- World Health Organization
XGBoost- Extreme Gradient Boosting