مسئولیت متن و شکل بدوش نویسندهٔ مضمون میباشد، عقیدهٔ نویسنده لزوماً نظر افغان جرمن آنلاین نمی باشد

لیکوال: دوکتور مسعود پښتون
رالېږونکی: اتل بنوال

۲۰۲۶/۰۷/۲۱

د محاسبوي ژبپوهنې پېژندنه،

نظرياتي بنسټونه، علمي لیدلوري او په اوسني عصر کې یې اړتیا

محاسبوي ژبپوهنه (Computational Linguistics) د ژبپوهنې، کمپیوټر ساینس، ریاضیاتو، احصایي، ادراکي علومو(Cognitive Science) ، مصنوعي ځیرکتیا (Artificial Intelligence) او منطق د ګډو هڅو پایله ده. د دې علم بنسټ پر دې ولاړ دی چې ایا د انسان ژبه په داسې علمي او محاسبوي چوکاټ کې تشریح کېدای شي چې کمپیوټر يې درک، تحلیل او تولید کړای شي؟ همدې پوښتنې د محاسبوي ژبپوهنې د نظریاتو د رامنځته کېدو لامل شوې ده.
د دې علم پرمختګ یوازې د کمپیوټرونو د چټکتیا محصول نه دی، بلکې د هغو ژبپوهانو، فیلسوفانو او کمپیوټرپوهانو د نظریاتو پایله هم ده چې د انسان د ژبې د جوړښت، مانا او ادراک په اړه يې بنسټيزې څېړنې کړې دي. د دغو نظریاتو په رڼا کې نن ورځ داسې هوښیار ژبني ماډلونه جوړ شوي چې د میلیونونو انسانانو د ورځني ژوند برخه ګرځېدلي دي.

د محاسبوي ژبپوهنې (Computational Linguistics) تاریخ:

محاسبوي ژبپوهنه (Computational Linguistics) د ژبپوهنې، کمپیوټر ساینس، مصنوعي ځیرکتیا (Artificial Intelligence)، ریاضیاتو او ادراکي علومو (Cognitive Science) ترمنځ یوه ګډه او بین‌الرشتوي علمي څانګه ده، چې موخه یې د انساني ژبې د تحلیل، درک، تولید او ژباړې لپاره د کمپیوټري سیستمونو جوړول او پراختیا ده. نن ورځ دا علم د مصنوعي ځیرکتیا او طبیعي ژبې د پروسس (Natural Language Processing) له مهمو بنسټونو څخه شمېرل کېږي، خو د دې علم تاریخ تر اوو لسیزو زیات عمر لري او د بېلابېلو پړاوونو له تېرېدو وروسته یې اوسنی پرمختللی بڼه خپله کړې ده.
د محاسبوي ژبپوهنې بنسټ په رسمي ډول د شلمې پېړۍ د پنځوسمو کلونو له پیل څخه اېښودل شوی، خو فکري بنسټ يې په ۱۹۴۹ زېږدیز کال کې د امریکایي ریاضي‌پوه او ساینس‌پوه وارن ویور (Warren Weaver) له خوا کېښودل شو. نوموړي د Translation په نوم یو مشهور یادښت (Memorandum) خپور کړ او پکې یې دا نظریه وړاندې کړه چې کمپیوټرونه د ریاضیاتي او احصایوي اصولو په مرسته د بېلابېلو ژبو ترمنځ ژباړه ترسره کولای شي. ویور باور درلود چې د انسان ژبې هم د رمزونو (Codes) په څېر ځانګړي جوړښتونه لري او کمپیوټرونه يې تحلیل او ژباړه کولای شي. له همدې امله ډېری څېړونکي وارن ویور د محاسبوي ژبپوهنې د فکري بنسټ اېښودونکي په توګه یادوي.
د ویور له نظریې وروسته په متحده ایالاتو کې ګڼو پوهنتونونو او څېړنیزو مرکزونو د ماشیني ژباړې (Machine Translation)په اړه عملي څېړنې پیل کړې. په ځانګړې توګه د جورجټاون پوهنتون (Georgetown University)او د IBM شرکت ګډه پروژه د دې علم په تاریخ کې یو مهم پړاو ګڼل کېږي.
په ۱۹۵۴ زېږدیز کال کې د Georgetown–IBM Experiment په نوم ازمايښت ترسره شو چې پکې کمپیوټر شاوخوا شپېته روسي جملې په انګلیسي ژبه وژباړلې. که څه هم دغه ژباړه د محدودو لغتونو او ګرامري قواعدو پر

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په خیر و لولئ

بنسټ وه، خو دا لومړی ځل و چې نړۍ په عملي ډول ولیدل چې کمپیوټر د انسان ژبه پروسس او ژباړلای شي.
همدې بریا د نړۍ د ډېرو هېوادونو پام دې نوې علمي څانګې ته واړاوه.
د ۱۹۵۰ او ۱۹۶۰ لسیزو په اوږدو کې متحده ایالات د محاسبوي ژبپوهنې د څېړنو نړیوال مرکز وګرځېد. د سړې جګړې (Cold War) پر مهال د روسي علمي او پوځي اسنادو ژباړې ته اړتیا د دې لامل شوه چې د امریکا حکومت په ځانګړې توګه د دفاع وزارت او نورو علمي بنسټونو په دې برخه کې پراخه مالي پانګونه وکړي. په همدې موده کې څېړنې تر ډېره د قاعده‌محورو (Rule-Based) سیستمونو پر جوړولو، صرفي تحلیل (Morphological Analysis)، نحوي تحلیل (Syntactic Parsing) او د لغتونو د بریښنایي قاموسونو پر جوړولو متمرکزې وې.
په همدې لسیزه کې د امریکایي ژبپوه نوم چامسکي (Noam Chomsky) د بدلونيزې - زېږندې ګرامر (Transformational-Generative Grammar)نظریه هم وړاندې شوه. چامسکي په ۱۹۵۷ کال کې خپل مشهور کتاب Syntactic Structures خپور کړ چې د ژبې د نحوي جوړښتونو علمي تحلیل یې وړاندې کاوه.
که څه هم چامسکي خپله د محاسبوي ژبپوهنې بنسټ ایښودونکی نه و، خو د هغه نحوي نظریې د محاسبوي ژبپوهنې د ډېرو لومړنیو سیستمونو لپاره مهم نظري بنسټ برابر کړ او د کمپیوټري ګرامرونو په جوړولو کې ترې پراخه ګټه واخیستل شوه.
سره له دې چې په لومړیو کلونو کې د ماشیني ژباړې په اړه ډېرې هیلې موجودې وې، خو عملي پایلې د تمې خلاف وې. په ۱۹۶۶ کال کې د امریکا د حکومت له خوا د ALPAC (Automatic Language Processing Advisory Committee)په نوم کمېټې خپل راپور خپور کړ. راپور دې پایلې ته ورسېد چې ماشیني ژباړه لا تر اوسه د عملي استفادې وړ نه ده او د ژباړې کیفیت یې کمزوری دی. د همدې راپور له خپرېدو وروسته د امریکا حکومت د ماشیني ژباړې ډېری پروژې ودرولې او مالي ملاتړ یې کم کړ. دا پړاو د محاسبوي ژبپوهنې په تاریخ کې د لومړي ستر رکود په نوم یادېږي.
سره له دې چې په امریکا کې څېړنې کمزورې شوې، خو په ۱۹۷۰ لسیزه کې په بریتانیا، فرانسه، جرمني، شوروي اتحاد، جاپان او کاناډا کې د محاسبوي ژبپوهنې څېړنې دوام پیدا کړ. په دې هېوادونو کې پوهنتونونو ځانګړي څېړنیز مرکزونه جوړ کړل او د ژبپوهانو او کمپیوټرپوهانو ګډ کار زیات شو. په همدې موده کې څېړنې یوازې په ماشیني ژباړه محدودې نه وې، بلکې د وینا پېژندنې(Speech Recognition) ، معلوماتو راایستلو (Information Retrieval)، ګرامري سمون، او د متن د تحلیل په برخو کې هم پراختیا ومونده.
د ۱۹۸۰ لسیزه د محاسبوي ژبپوهنې لپاره د بدلون لسیزه وه. څېړونکي ورو ورو د قاعده‌محورو سیستمونو پر ځای احصایوي (Statistical) طریقو ته مخه کړه. د کمپیوټرونو د حافظې او محاسبوي وړتیاوو له زیاتېدو سره د ژبو ستر بریښنایي متنونه یا کورپسونه (Corpora) جوړ شول. د امریکا Brown Corpus او د بریتانیا British National Corpus (BNC)د نړۍ له لومړنیو لویو ژبنیو کورپسونو څخه وو، چې څېړونکو ته یې د ژبې د واقعي استعمال د تحلیل زمینه برابره کړه. په همدې لسیزه کې د وینا پېژندنې، اتومات متن‌تحلیل او احصایوي نحوي تحلیل په برخو کې هم د پام وړ پرمختګ وشو.
په ۱۹۹۰ لسیزه کې د ماشیني زده‌کړې (Machine Learning) کارونه پیل شول. د دې پر ځای چې ټول قواعد په لاس ولیکل شي، کمپیوټرونو ته د بېلګو پر بنسټ زده‌کړه ورکول کېده. د همدې پرمختګ په مرسته د وینا برخو پېژندنه(Part-of-Speech Tagging) ، د نوم لرونکو موجودیتونو پېژندنه (Named Entity Recognition)، نحوي تحلیل او د معلوماتو استخراج په بې‌ساري ډول پرمختګ وکړ. په همدې وخت کې د محاسبوي ژبپوهنې نړیواله ټولنه(Association for Computational Linguistics - ACL) ، چې په ۱۹۶۲ کال کې جوړه شوې وه، د نړۍ د دې علم تر ټولو مهمه علمي اداره وګرځېده او تر نن ورځې پورې د نړۍ ستر نړیوال کنفرانسونه او علمي مجلې خپروي.
له ۲۰۰۰ زېږدیز کال وروسته د انټرنېټ پراختیا د محاسبوي ژبپوهنې پرمختګ چټک کړ. د انټرنېټ له لارې د میلیاردونو متنونو راټولېدل، د ګوګل، یاهو او نورو لټون‌انجنونو رامنځته کېدل، او د څو ژبو موازي متنونو (Parallel Corpora)زیاتوالی د ماشیني ژباړې کیفیت ډېر ښه کړ. دIBM Research ، Google، Microsoftاو نورو نړیوالو ټکنالوژۍ شرکتونو د احصایوي ماشیني ژباړې (Statistical Machine Translation)په برخه کې ستري لاسته راوړنې درلودې.
د ۲۰۱۰ لسیزه د ژورې زده‌کړې (Deep Learning) او عصبي شبکو (Neural Networks) د پیل لسیزه وه.
په ۲۰۱۳ کال کې د Word2Vec ماډل د کلمو مانیزې اړیکې په ریاضیاتي ډول وښودلې. وروسته د Recurrent Neural Networks (RNNs)او Long Short-Term Memory (LSTM) ماډلونو د ژبې د تحلیل او ژباړې دقت زیات کړ.
تر ټولو ستر بدلون په ۲۰۱۷ کال کې رامنځته شو، کله چې د ګوګل څېړونکو د Transformer معمارۍ معرفي کړه او خپله مشهوره مقاله Attention Is All You Need یې خپره کړه. همدې نوښت د محاسبوي ژبپوهنې بڼه بدله کړه او د لویو ژبنیو ماډلونو (Large Language Models) بنسټ یې کېښود.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکني د لیکنیزې بنې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځير و لولئ

له ۲۰۱۸ کال وروسته دBERT ، GPT، RoBERTa، T5او نورو لویو ژبنیو ماډلونو په رامنځته کېدو سره محاسبوي ژبپوهنه د مصنوعي ځيرکتيا له مهمو برخو څخه وګرځېده. دغه ماډلونه د میلیاردونو کلمو پر بنسټ روزل شوي او د ژباړې، متن لنډوني، پوښتنو ځوابولو، احساساتي تحلیل، خبرو اترو او متن تولید په برخو کې د انسان کچې ته نږدې پایلې وړاندې کوي.
نن ورځ محاسبوي ژبپوهنه د نړۍ په ډېرو پرمختللو هېوادونو لکه متحده ایالات، کاناډا، بریتانیا، جرمني، فرانسه، هالنډ، فنلنډ، سویډن، چین، جاپان، سویلي کوریا، هند، سنګاپور او استرالیا کې د پوهنتونونو، څېړنیزو مرکزونو او ټکنالوژۍ د لویو شرکتونو له مهمو څېړنیزو برخو څخه ده. سربېره پر دې، د افغانستان، پاکستان، نیپال، افریقایي هېوادونو او نورو هغو هېوادونو لپاره هم څېړنې رواني دي چې د ټیټو ډیجیټلي سرچینو لرونکي ژبې (Low-Resource Languages)لري، ترڅو د دغو ژبو لپاره برېښنایي قاموسونه، ژبني کورپسونه، د وینا پېژندنې سیستمونه او ماشیني ژباړې پروګرامونه جوړ شي.
په ټوله کې، د محاسبوي ژبپوهنې تاریخ له ۱۹۴۹ کال کې د وارن ویور له نظريي څخه پیل، د ۱۹۵۴ کال د جورجټاون IBM—له ازمایښت څخه تر ننني لويو ژبنیو ماډلونو پورې د دوامداره علمي پرمختګ کیسه ده. دا علم د قاعده‌محورو سیستمونو، احصایوي میتودونو، ماشیني زده‌کړې، ژورې زده‌کړې او عصبي شبکو له پړاوونو تېر شوی او نن د مصنوعي ځيرکتيا له تر ټولو اغېزمنو او چټک پرمختګ کوونکو علمي برخو څخه ګڼل کېږي. د دې برخې راتلونکې لا هم د پراخو نوښتونو، د نړۍ د ټولو ژبو د ډیجیټلي کولو او د انسان او ماشین ترمنځ د لا طبیعي اړیکو د رامنځته کولو ژمنه کوي.

ژبه د یوه نظام په توګه:

د معاصرې ژبپوهنې بنسټیز فکر دا دی چې ژبه د تصادفي کلمو ټولګه نه ده، بلکې یو منظم او قانونمند نظام دی. د همدې اصل پر بنسټ محاسبوي ژبپوهنه هڅه کوي چې د ژبې دا نظام په محاسبوي فورمولونو، الګوریتمونو او ماډلونو بدل کړي.
د ژبې هره کچه—له غږ څخه نیولې تر مانا او ګفتمان پورې—خپل ځانګړي اصول لري. که دا اصول په دقیق ډول وپېژندل شي، نو کمپیوټر هم کولای شي د انسان ژبه تحلیل کړي. د همدې لیدلوري له مخې، ژبه د معلوماتو یو جوړښتیز سیستم ګڼل کېږي چې د محاسبوي تحلیل وړتیا لري.

جوړښتي ژبپوهنه او محاسبوي ژبپوهنې بنسټ:

د شلمې پېړۍ په پیل کې د سویس ژبپوه فردیناند دو سوسور د جوړښتي ژبپوهنې (Structural Linguistics) نظریه وړاندې کړه. هغه ژبه د نښو (Signs) یو منظم نظام وباله او څرګنده یې کړه چې د ژبې هره نښه له دوو برخو جوړه ده: دال (Signifier) او مدلول.(Signified)
د سوسور له نظره د ژبې ارزښت د کلمو په خپلواک وجود کې نه، بلکې د هغوی په متقابلو اړیکو کې نغښتی دی. همدې نظريي وروسته د محاسبوي ژبپوهنې لپاره لاره هواره کړه، ځکه کمپیوټر هم د اړیکو، شبکو او نمونو له لارې معلومات تحلیلوي.
نن ورځ د کلمو د اړیکو د تحلیل لپاره د Word Embeddings ، Semantic Networks او Knowledge Graphs په څېر پرمختللي محاسبوي تخنیکونه کارول کېږي، چې فکري ریښه یې د جوړښتي ژبپوهنې له همدې نظریې سره تړاو لري.

د نوام چامسکي د تولیدي ګرامر نظریه:

د محاسبوي ژبپوهنې په تاریخ کې د نوام چامسکي نظریات تر ټولو اغېزمن بلل کېږي. چامسکي په ۱۹۵۷ کال کې د تولیدي ګرامر (Generative Grammar) نظریه وړاندې کړه او ویې ویل چې انسان یوازې جملې نه حفظ کوي، بلکې په ذهن کې داسې قواعد لري چې له هغوی څخه بې شمېره نوې جملې تولیدولای شي.
چامسکي باور درلود چې ژبه د انسان د ذهن یوه فطري وړتیا ده. د هغه په اند هر انسان د ژبې د زده کړې لپاره له زېږون سره یو ذهني چوکاټ لري چې د نړیوال ګرامر (Universal Grammar) په نوم یادېږي.
د دې نظریې اغېز پر محاسبوي ژبپوهنه ډېر ژور و، ځکه د لومړنیو نحوي تحلیلګرونو (Parsers) ډېرې الګوریتمونه د همدې تولیدي قواعدو پر بنسټ جوړ شول. د Context-Free Grammar ، Phrase Structure Grammar او Parsing Algorithms په پراختیا کې د چامسکي نظریه بنسټیز رول لري. سره له

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

دي، وروسته څېړنو وښودله چې یوازې نحوي قواعد د ژبې د بشپړ درک لپاره بسنه نه کوي، ځکه ژبه د سیاق، مانا او ټولنیزو شرایطو تر اغېز لاندې هم بدلېږي.

د الن ټورینګ نظریه:

د کمپیوټر ساینس مخکښ الن ټورینګ په ۱۹۵۰ م کال کې خپله مشهوره مقاله Computing Machinery and Intelligenceخپره کړه او پکې یې دا پوښتنه مطرح کړه: «ایا ماشین فکر کولای شي؟ هغه د ټورینګ ازموینه (Turing Test)وړاندې کړه. د دې ازموینې له مخې، که یو کمپیوټر وکولای شي د خبرو اترو پر مهال انسان په دې قانع کړي چې مقابل لوری هم انسان دی، نو ویل کېدای شي چې ماشین هوښیار چلند کوي.
د ټورینګ نظریه د محاسبوي ژبپوهنې لپاره تاریخي اهمیت لري، ځکه دا یې څرګنده کړه چې طبیعي ژبه د مصنوعي ځیرکتیا د ارزونې تر ټولو مهم معیار دی. همدا نظریه د ننني چټبوټونو او لویو ژبنیو ماډلونو د ارزونې لپاره هم الهامبښونکې ده.

ادراکي ژبپوهنه او محاسبوي ماډلونه:

په ۱۹۸۰ م لسیزه کې د ادراکي ژبپوهنې (Cognitive Linguistics) نظریې پراخې شوې. د دې مکتب استازي استدلال کوي چې ژبه د انسان له ادراک، تجربې او مفهومي نظام څخه جلا نه ده. یعني د یوې کلمې مانا یوازې د قاموس له تعریف څخه نه، بلکې د انسان له ذهني تجربو، فرهنګي شالید او سیاق څخه هم اغېزمنه کېږي.
دغه لیدلوری د محاسبوي ژبپوهنې لپاره نوی ځلنج و. پخوانیو قاعده‌محورو سیستمونو یوازې نحوي جوړښت ته پام کاوه، خو د ادراکي ژبپوهنې نظریې وښودله چې د ژبې مانا د انساني تجربې له درک پرته بشپړه نه شي تشریح کېدای. د همدې اړتیا له امله اوسني ژبني ماډلونه هڅه کوي چې د کلمو ماناوې د هغوی د استعمال له سیاق سره یوځای زده کړي. د Contextual Word Representations او Transformer معمارۍ د همدې مفکورې پرمختللې بېلګې دي.

احصایوي ژبپوهنه:

د شلمې پېړۍ په وروستیو لسیزو کې د ژبې د تحلیل لپاره احصایوي میتودونه ډېر مشهور شول. د دې نظریې له مخې، ژبه د احتمالي اړیکو یو نظام دی او د لویو متني زېرمو (Corpora) د تحلیل له لارې یې نمونې موندل کېدای شي. په دې پړاو کې د N-gram ماډلونه، Hidden Markov Modelsاو Bayesian Models رامنځته شول. دغو ماډلونو ماشین ته دا توان ورکړ چې د کلمو د تکرار، ګډ راتګ او احتمالي اړیکو پر بنسټ راتلونکې کلمه اټکل کړي. احصایوي ژبپوهنې د ماشین ژباړې، غږ پېژندنې او املا سمونې په برخه کې د پام وړ پرمختګونه رامنځته کړل، خو د ژبې ژوره مانا یې لا هم په بشپړه توګه نه شوه درک کولای.

ژوره زده کړه او عصبي ژبني ماډلونه:

له ۲۰۱۰ م کال وروسته د ژورو عصبي شبکو (Deep Neural Networks) پراختیا د محاسبوي ژبپوهنې په تاریخ کې یو نوی فصل پرانیست. په دې پړاو کې کمپیوټر نور یوازې قواعد یا احصایوي اړیکې نه زده کوي، بلکې د میلیونونو متنونو له تحلیل څخه ژورې بڼې او اړیکې هم کشفوي. عصبي ژبني ماډلونه د انسان د عصبي سیستم له جوړښت څخه الهام اخلي. دغه ماډلونه په سلګونو یا زرګونو پټ پوړونه لري چې هر یو یې د ژبې بېلابېل ځانګړتیاوې زده کوي. د همدې پرمختګ له امله نن ورځ کمپیوټر د متن لنډیز، پوښتنو ته ځواب، ژباړه، متن تولید او حتی علمي لیکنې هم ترسره کولای شي.

د ترانسفارمر معمارۍ انقلاب:

په ۲۰۱۷ م کال کې د Transformer معمارۍ معرفي کېدل د محاسبوي ژبپوهنې په تاریخ کې یو ستر بدلون و. د دې معمارۍ تر ټولو مهم نوښت د Attention Mechanism و، چې ماډل ته اجازه ورکوي د جملې ټولې برخې په یو وخت کې وارزوي او د کلمو ترمنځ لرې اړیکې هم درک کړي. د همدې تخنیک پر بنسټ وروسته لوی ژبني ماډلونه (Large Language Models) جوړ شول چې په میلیاردونو کلمو روزل شوي دي. دغه ماډلونه نه یوازې د ژبې نحوي جوړښت، بلکې د مانا، سیاق او ګفتمان ډېرې اړخونه هم زده کوي.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

د معناپوهنې او سیاق اهمیت:

د ژبې تر ټولو ستونزمن اړخ مانا (Semantics) او سیاق (Context) دی. یوه کلمه په بېلابېلو حالاتو کې بېلابېلې ماناوې لرلای شي. د بېلګې په توګه، د «سر» کلمه کله د بدن غړی، کله مشر او کله د کوم شي پیل ته اشاره کوي. د دې توپیر پېژندل د محاسبوي ژبپوهنې له مهمو ننګونو څخه دي. اوسني ژبني ماډلونه هڅه کوي چې د جملې ټول سیاق تحلیل کړي او د کلمې مناسبه مانا وټاکي. دا پرمختګ د ژبې د طبیعي درک په لور یو مهم ګام ګڼل کېږي.

د نظریاتو ترمنځ اړیکه:

که د محاسبوي ژبپوهنې تاریخ ته وکتل شي، څرګندېږي چې دا علم د یوې نظریې محصول نه دی. جوړښتي ژبپوهنې د ژبې نظام تشریح کړ، تولیدي ګرامر د نحوي قواعدو اهمیت روښانه کړ، ادراکي ژبپوهنې د انسان د ذهن رول بیان کړ، احصایوي میتودونو د معلوماتو له لارې زده کړه ممکنه کړه او ژورو عصبي شبکو د دغو ټولو نظریاتو عملي پلي کېدو ته نوې لار هواره کړه.
له همدې امله، معاصر محاسبوي ژبپوهنه د ژبپوهنې، احصایې، ریاضیاتو، کمپیوټر ساینس، مصنوعي ځیرکتیا او ادراکي علومو د ګډو نظریو یوه علمي مجموعه ده. د دې علم راتلونکې هم په همدې بینرشتوي همکارۍ پورې تړلې دی، ځکه د انسان د ژبې بشپړ درک یوازې د یوې علمي څانګې له لارې ممکن نه دی.
طبیعي ژبې پروسس(Natural Language Processing) ، ماشیني زده کړه او لوی ژبني ماډلونه(LLMs) د شلمې پېړۍ په وروستیو او په ځانګړې توګه د یوویشتمې پېړۍ په لومړیو لسیزو کې د کمپیوټر ساینس او مصنوعي ځیرکتیا چټک پرمختګ د دې لامل شو چې د انسان او ماشین ترمنځ اړیکې نوې بڼه خپله کړي. که پخوا کمپیوټر یوازې عددي محاسبې او ټاکلي برنامې اجرا کولې، نن ورځ کولای شي متن ولولي، خبرې واوري، پوښتنو ته ځوابونه ورکړي، مقالې ولیکي، ژباړه وکړي او حتی د انسان په څېر منطقي استدلال وړاندې کړي. د دې ټولو پرمختګونو تر شا اساسي علمي بنسټ طبیعي ژبې پروسس (Natural Language Processing – NLP) دی، چې د محاسبوي ژبپوهنې تر ټولو مهمه او عملي څانګه ګڼل کېږي.
طبیعي ژبې پروسس هغه علمي او تخنیکي ډګر دی چې موخه یې دا ده چې کمپیوټر د انسان طبیعي ژبه درک، تحلیل، تفسیر او تولید کړي. دلته د "طبیعي ژبې" څخه موخه هغه ژبې دي چې انسانان یې په ورځني ژوند کې کاروي، لکه پښتو، دري، انګلیسي، عربي، فرانسوي او نورې ژبې. د NLP اساسي موخه دا ده چې ماشین د انسان د ژبې له ظاهري جوړښت څخه تېر شي او د هغې مانا، سیاق، نیت او اړیکې درک کړي.
د طبیعي ژبې پروسس د محاسبوي ژبپوهنې او مصنوعي ځیرکتیا ترمنځ د وصل پُل دی. محاسبوي ژبپوهنه د ژبې علمي او نظریاتي اصول وړاندې کوي، په داسې حال کې چې NLP همدغه اصول په عملي الګوریتمونو او کمپیوټري سیستمونو کې پلي کوي. په همدې دلیل، د دواړو اصطلاحاتو ترمنځ ډېر نږدې اړیکه شته، که څه هم محاسبوي ژبپوهنه تر ډېره علمي او تیوریکه بڼه لري او NLP عملي او انجنیرۍ اړخ استازیتوب کوي.

د طبیعي ژبې تاریخي پروسس او پرمختګ:

د طبیعي ژبې پروسس لومړني ګامونه د ۱۹۵۰ او ۱۹۶۰ لسیزو په جریان کې پورته شول، کله چې څېړونکو هڅه وکړه د ماشین ژباړې لومړني سیستمونه جوړ کړي. په هغه وخت کې ټول سیستمونه د قواعدو (Rules) پر بنسټ وو. ژبپوهانو به د یوې ژبې نحوي، صرفي او لغوي قواعد په لاسي ډول لیکل او کمپیوټر به د هماغو قواعدو له مخې متن تحلیل کاوه. که څه هم دغه سیستمونه د محدودو متنونو لپاره اغېزمن وو، خو د انسان د ژبې د انعطاف، ابهام او استثناوو له امله یې دقت کم و. د بېلګې په توګه، یوه کلمه په بېلابېلو سیاقونو کې بېلابېلې ماناوې لري، خو قاعده‌محور سیستمونه دا توپیر نه شوای تشخیص کولای.
په ۱۹۹۰ م لسیزه کې د احصایوي ژبپوهنې (Statistical NLP) دور پیل شو. په دې پړاو کې کمپیوټر نور یوازې قواعد نه تعقیبول، بلکې د میلیونونو جملو له تحلیل څخه نمونې زده کوي. د همدې بدلون له امله د ژباړې، متن طبقه‌بندۍ او وینا پېژندنې سیستمونه د پام وړ ښه شول. وروسته د ژورو عصبي شبکو (Deep Neural Networks)او په ځانګړې توګه د Transformer معمارۍ راڅرګندېدو د NLP په تاریخ کې ستر انقلاب رامنځته کړ. اوسني ژبني ماډلونه د میلیاردونو کلمو له تحلیل وروسته د ژبې داسې ژورې اړیکې زده کوي چې پخوا یې تصور هم نه کېده.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

د ژبې د طبیعي پروسس مهمي دندې:

د طبیعي ژبې پروسس بېلابېلي علمي او عملي دندې لري، چې هره یوه یې د ژبې د درک ځانګړی اړخ بیانوي.
لومړی، د متن وېش (Tokenization) دی. په دې پړاو کې متن په جملو، کلمو یا کوچنیو واحدونو وېشل کېږي.
دا د NLP د ټولو راتلونکو مرحلو بنسټ ګڼل کېږي.
دویم، صرفي تحلیل (Morphological Analysis) دی. دلته کمپیوټر د کلمې رېښه، مخکیني او وروستي ترړنګونه (Prefixes)او (Suffixes) او ګرامري ځانګړتیاوې تشخیصوي. دا پړاو د هغو ژبو لپاره ډېر مهم دی چې صرفي جوړښت یې پېچلی وي، لکه پښتو.
درېیم، نحوي تحلیل (Parsing) دی. په دې مرحله کې کمپیوټر د جملې نحوي جوړښت تحلیلوي او د کلمو ترمنځ اړیکې معلوموي.
څلورم، ماناپوهنیز تحلیل (Semantic Analysis) دی. په دې پړاو کې ماشین هڅه کوي چې د جملې حقیقي مانا درک کړي، نه یوازې د کلمو ظاهري ترتیب.
پنځم، عملي پوهنیز تحلیل (c Analysis) دی. دلته د جملې مانا د ټولنیز او سیاقي وضعیت له مخې ارزول کېږي.
د بېلګې په توګه، د «دروازه خلاصه ده» جمله کېدای شي یوازې خبر وي یا په غیر مستقیم ډول د دروازې د بندولو غوښتنه هم وي.

ماشیني زده کړه او د ژبنی تحلیل:

ماشین زده کړه (Machine Learning) د مصنوعي ځیرکتیا هغه څانګه ده چې کمپیوټر ته اجازه ورکوي د معلوماتو له تحلیل څخه زده کړه وکړي، پرته له دې چې هره قاعده ورته په لاسي ډول ولیکل شي. په محاسبوي ژبپوهنه کې ماشین زده کړه د دې لپاره کارول کېږي چې کمپیوټر د میلیونونو متنونو له لوستلو وروسته د ژبې قواعد، اړیکې او احتمالي نمونې پخپله زده کړي. د بېلګې په توګه، که یو سیستم میلیونونه پښتو جملې تحلیل کړي، ورو ورو زده کوي چې کومې کلمې عموماً یو ځای راځي، کومې نحوي جوړښتونه ډېر کارېږي او کومې ماناوې له کوم سیاق سره تړاو لري.

ماشین زده کړه په درې اساسي ډولونو وېشل کېږي:

۱ .څارل شوې زده کړه(Supervised Learning) ، چې پکې سیستم د نښه‌لرونکو معلوماتو پر بنسټ روزل کېږي.
۲ .ناڅارل شوې زده کړه(Unsupervised Learning) ، چې پکې ماشین پخپله د معلوماتو پټې نمونې کشفوي.
۳ .پیاوړې زده کړه(Reinforcement Learning) ، چې پکې سیستم د انعام او سزا له لارې د غوره پرېکړو زده کړه کوي.
په معاصرو ژبنیو ماډلونو کې د دغو ټولو میتودونو بېلابېل ترکیبونه کارول کېږي.
ژوره زده کړه(Deep Learning)
ژوره زده کړه د ماشین زده کړې پرمختللې بڼه ده چې د مصنوعي عصبي شبکو (Artificial Neural Networks)پر بنسټ ولاړه ده. دغه شبکې د انسان د مغز له عصبي جوړښت څخه الهام اخلي. په دودیزو ماشیني سیستمونو کې انسان اړ و چې مهمې ځانګړتیاوې (Features) پخپله وټاکي، خو په ژورو زده کړو کې سیستم دا ځانګړتیاوې پخپله کشفوي. د همدې ځانګړتیا له امله ژوره زده کړه د ژبې د درک په برخه کې بې ساري پرمختګونه رامنځته کړل. د بېلګې په توګه، که یو ژبنی ماډل په میلیاردونو جملو وروزل شي، هغه نه یوازې د کلمو اړیکې زده کوي، بلکې د سبک، استدلال، سیاق، احساساتو او حتی د لیکوال د بیان ځانګړتیاوې هم درک کولای شي.

د ټرانسفارمر معمار:(Transformer Architecture)

په ۲۰۱۷ کال کې د "Attention Is All You Need" تر عنوان لاندې څېړنیزې مقالې د ژبنیو ماډلونو په تاریخ کې نوی باب پرانیست. په دې مقاله کې د Transformer معمارۍ معرفي شوه. د دې معمارۍ تر ټولو مهم نوښت د Self-Attention Mechanism دی. دا تخنیک ماډل ته اجازه ورکوي چې د جملې ټولې کلمې په یو وخت کې تحلیل کړي او د هغوی ترمنځ اړیکې وپېژني. د پخوانیو تکراري شبکو (RNNs) برعکس، ټرانسفارمر کولای شي اوږده متنونه په ډېر لوړ سرعت او دقت تحلیل کړي. د همدې معمارۍ پر بنسټ د نن ورځې لوی ژبني ماډلونه جوړ شوي دي.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له راليږلو مخکې په خیر و لولئ

لوی ژبني ماډلونه:(Large Language Models)

لوی ژبني ماډلونه یا LLMs د مصنوعي ځیرکتیا هغه پرمختللي سیستمونه دي چې د میلیاردونو کلمو او اسنادو له تحلیل وروسته روزل کېږي. دغه ماډلونه د ژبې د تولید، تحلیل، لنډیز، ژباړې، استدلال او پوښتنو ځوابولو وړتیا لري. د دې ماډلونو روزنه په څو پړاوونو کې ترسره کېږي. لومړی، سیستم د پراخو متني زېرمو پر بنسټ د ژبې عمومي جوړښت زده کوي. وروسته د ځانګړو دندو لپاره نوره روزنه ورکول کېږي. په وروستي پړاو کې د انساني ارزوونکو د فیډبک له لارې د ځوابونو کیفیت ښه کېږي. دغه ماډلونه د دې توان لري چې د متن له سیاق څخه د راتلونکي کلمې اټکل وکړي، خو د روزنې د پراخوالي له امله یې ځوابونه د انساني استدلال په څېر ښکاري.

د NLP او LLMs عملي اغېزې:

د طبیعي ژبې پروسس او لویو ژبنیو ماډلونو پرمختګ د ژوند په نږدې ټولو برخو خپلې اغېزې لري. په تعلیم کې د هوښیارو ښوونیزو مرستیالانو، په روغتیا کې د طبي اسنادو د تحلیل، په حقوقو کې د اسنادو د ارزونې، په سوداګرۍ کې د پیرودونکو د خدمتونو، په رسنیو کې د خبرونو د لنډیز، او په څېړنه کې د علمي معلوماتو د تحلیل لپاره دا ټکنالوژي په پراخه کچه کارول کېږي.
په ورته وخت کې، د ټیټو سرچینو ژبو لکه پښتو لپاره هم دا پرمختګونه نوی فرصت برابروي. که د پښتو لپاره معیاري متني زېرمې، لغتنامې او ژبني ډیټاسیټونه جوړ شي، نو د دې ژبې لپاره هم د لوړ کیفیت ژبني ماډلونه رامنځته کېدای شي. د طبیعي ژبې پروسس او لوی ژبني ماډلونه د محاسبوي ژبپوهنې تر ټولو مهمه لاسته راوړنه ده. د دې ټکنالوژۍ د چټک پرمختګ له امله انسان او ماشین ورځ تر بلې طبیعي، اغېزمنه او هوښیاره اړیکه رامنځته کوي. تمه کېږي چې په راتلونکو کلونو کې به د NLP او LLMs پرمختګ د ژبې، زده کړې، څېړنې او ټولنیزو اړیکو بڼه لا ډېره بدله کړي.
د محاسبوي ژبپوهنې عملي تطبیق: ماشیني ژباړه، غږ پېژندنه، متن تولید، احساساتو تحلیل، د معلوماتو رااېستل او هوښیار چټبوټونه
د محاسبوي ژبپوهنې ارزښت یوازې په دې کې نه دی چې د ژبې په اړه نوې علمي نظریې وړاندې کوي، بلکې تر ټولو مهمه ځانګړنه یې دا ده چې دغه نظریې په عملي ټکنالوژیو بدلېږي. نن ورځ د نړۍ ډېری هوښیار ډیجیټلي خدمتونه د محاسبوي ژبپوهنې پر بنسټ ولاړ دي. د طبیعي ژبې پروسس، ماشین زده کړې او ژورو عصبي شبکو په پرمختګ سره کمپیوټرونه توانېدلي چې د انسان ژبه په داسې کچه درک او تولید کړي چې څو لسیزې وړاندې یې تصور هم ناشونی ګڼل کېده. له همدې امله، محاسبوي ژبپوهنه اوس د روغتیا، زده کړې، سوداګرۍ، رسنیو، امنیت، حکومتولۍ، حقوقو او علمي څېړنو په ګډون د ژوند په نږدې ټولو برخو کې اساسي رول لري.

ماشیني ژباړه:(Machine Translation)

ماشین ژباړه د محاسبوي ژبپوهنې له تر ټولو پخوانیو او اغېزناکو تطبیقونو څخه ده. موخه یې دا ده چې یو متن له یوې ژبې څخه بلې ژبې ته په اتومات ډول وژباړي، پرته له دې چې انسان مستقیمه ونډه ولري. د ماشین ژباړې لومړني سیستمونه د قواعدو پر بنسټ جوړ شوي وو او د ژبپوهانو له لوري په لاسي ډول ورته زرګونه ګرامري قواعد لیکل کېدل. دا سیستمونه د محدودو متنونو لپاره مناسب وو، خو د ژبې د ابهام، اصطلاحاتو او سیاق د درک وړتیا یې کمه وه. وروسته احصایوي ماشین ژباړه رامنځته شوه، چې د لویو موازي متني زېرمو (Parallel Corpora)پر بنسټ یې د ژبو ترمنځ احتمالي اړیکي زده کولې. په وروستیو کلونو کې عصبي ماشین ژباړې (Neural Machine Translation)د ژباړې کیفیت په بې‌ساري ډول لوړ کړی دی. دا سیستمونه د ټرانسفارمر معمارۍ او ژورو عصبي شبکو له لارې د متن سیاق، مانا او ګرامري جوړښت په یو وخت کې تحلیلوي او روانه ژباړه وړاندې کوي.
د پښتو ژبې لپاره د ماشین ژباړې پرمختګ لا هم له ننګونو سره مخ دی، ځکه د لوړ کیفیت موازي متني زېرمې محدودې دي. د معیاري ژبنیو معلوماتو راټولول، د اصطلاحاتو یو شان کول او د بېلابېلو لهجو ترمنځ همغږي د دې برخې له مهمو اړتیاوو څخه دي. که دا ستونزې حل شي، نو د پښتو لپاره به هم د لوړ کیفیت ژباړنیز سیستمونه رامنځته شي.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بنې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

غږ پېژندنه:(Speech Recognition)

غږ پېژندنه یا (ASR) Automatic Speech Recognition هغه ټکنالوژي ده، چې د انسان وینا په لیکلي متن بدلوي. دا ټکنالوژي د محاسبوي ژبپوهنې، غږپوهنې، ماشین زده کړې او سیګنال پروسس ګډه لاسته راوړنه ده. په دې سیستم کې لومړی د انسان غږ ثبتیږي، وروسته د غږ څپې په کوچنیو برخو وېشل کېږي، بیا فونیمونه، کلمې او جملې پېژندل کېږي او په پای کې متن تولیدېږي.
غږ پېژندنه په هوښیارو ټیلیفونونو، مجازي مرستیالانو، موټرو، روغتیايي وسایلو او د لاس‌ازاد خدمتونو په سیستمونو کې پراخه کارونه لري. د دې ټکنالوژۍ دقت د ژبې د غږیزو معلوماتو، لهجو او روزنیزو ډېټاسیټونو پر کیفیت پورې تړلی دی. د پښتو ژبې لپاره د غږ پېژندنې لویه ستونزه د معیاري غږیزو معلوماتو نشتوالی دی. د بېلابېلو سیمو لهجې، د غږونو توپیر او د ثبت شویو معلوماتو کمښت د دې سبب شوی چې د پښتو لپاره د غږ پېژندنې سیستمونه لا تر اوسه د نړیوالو ژبو په کچه پرمختګ ونه کړي.

د وینا تولید:(Text-to-Speech)

د غږ پېژندنې برعکس، د متن څخه وینا (Text-to-Speech) هغه ټکنالوژي ده چې لیکلی متن په طبیعي غږ بدلوي. دا سیستمونه لومړی متن تحلیلوي، د کلمو تلفظ ټاکي، د جملې اهنګ، وقفې او ټینګار محاسبه کوي او وروسته مصنوعي غږ تولیدوي. دغه ټکنالوژي د ړندو کسانو د مرستې، برېښنايي کتابونو، ښوونیزو پروګرامونو، ګرځنده اپلیکېشنونو او هوښیارو مرستیالانو لپاره ځانګړی اهمیت لري. د پښتو ژبې لپاره د طبیعي غږ تولید لا هم د پراخو څېړنو او معیاري غږیزو معلوماتو اړتیا لري.

د احساساتو تحلیل:(Sentiment Analysis)

احساساتو تحلیل د طبیعي ژبې پروسس هغه برخه ده چې د متن احساساتي محتوا تشخیصوي. په دې تحلیل کې ماشین هڅه کوي چې معلومه کړي یو متن مثبت، منفي او که بې‌طرفه احساس څرګندوي. دا ټکنالوژي په سوداګرۍ، بازارموندنه، ټولنیزو رسنیو، سیاست، عامه افکارو او د مشتریانو د رضایت د ارزونې لپاره پراخه کارول کېږي. د بېلګې په توګه، که زرګونه خلک د یوه محصول په اړه نظرونه ولیکي، سیستم کولای شي د هغو ټولو نظرونو عمومي احساس تحلیل کړي او معلومه کړي چې خلک تر ډېره بریده راضي دي که نه.
د احساساتو تحلیل یوازې د کلمو شمېرل نه دي، ځکه کله ناکله یوه جمله ظاهراً مثبته وي، خو په حقیقت کې طنز یا نیوکه څرګندوي. له همدې امله اوسني ژبني ماډلونه هڅه کوي چې د سیاق، سبک او ګفتمان له مخې احساسات تحلیل کړي.

د معلوماتو رااېستل:(Information Extraction)

د معلوماتو رااېستل د محاسبوي ژبپوهنې له مهمو تطبیقونو څخه دی چې موخه یې له اوږدو متنونو څخه د مهمو معلوماتو استخراج دی. په دې بهیر کې ماشین د اشخاصو، ځایونو، سازمانونو، نېټو، پېښو او نورو مهمو مفاهیمو پېژندنه کوي او وروسته یې په منظم ډول وړاندې کوي. دا ټکنالوژي په حقوقي اسنادو، طبي راپورونو، علمي مقالو، خبري رسنیو او دولتي معلوماتي سیستمونو کې ډېر ارزښت لري. د بېلګې په توګه، د زرګونو علمي مقالو له منځه د یوې ځانګړې ناروغۍ اړوند معلومات په څو ثانیو کې رااېستل کېدای شي.

متن لنډیز:(Automatic Text Summarization)

د معلوماتو د چټک زیاتوالي له امله د متن اتومات لنډیز ورځ تر بلې مهم کېږي. د دې ټکنالوژۍ موخه دا ده چې اوږده اسناد، راپورونه او مقالې په لنډ، خو مانا لرونکي متن بدل کړي. د متن لنډیز دوه ډوله دی. په لومړي ډول کې د متن مهمې جملې غوره کېږي، او په دویم ډول کې ماشین پخپله نوی لنډیز تولیدوي چې د اصلي متن مهمې مفکورې پکې خوندي وي. د ژورو عصبي شبکو او لویو ژبنیو ماډلونو پرمختګ د دویم ډول کیفیت ډېر لوړ کړی دی.
پوښتنو ته ځواب ویل:(Question Answering Systems)
د پوښتنو ځوابونکي سیستمونه هغه هوښیار پروګرامونه دي چې د کاروونکي پوښتنه تحلیلوي، د اړتیا وړ معلومات پیدا کوي او دقیق ځواب وړاندې کوي. دا سیستمونه په تعلیمي پلیټفارمونو، روغتیايي خدماتو، بانکدارۍ، حقوقي مشورو او علمي څېړنو کې پراخ کارول کېږي. د دې سیستمونو دقت د ژبې پر درک، د معلوماتو پر کیفیت او د

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بنې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په څېر و لولئ

استدلال پر ورتیا پورې تړلې دی. اوسني لوی ژبني ماډلونه د دې توان لري چې د پیچلو پوښتنو لپاره هم نسبي منطقي او منظم ځوابونه وړاندې کړي، خو بیا هم د دقیقو حقایقو د تایید اړتیا پاتې کیږي.

هوښیار چېټبوټونه:(Intelligent Chatbots)

هوښیار چېټبوټونه د محاسبوي ژبپوهنې او مصنوعي ځیرکتیا له تر ټولو مشهور تطبیقونو څخه دي. دا سیستمونه د انسان طبیعي ژبه درک کوي، د پوښتنې موخه تشخیصوي او د مناسب ځواب تولید هڅه کوي. د چېټبوټونو لومړني نسل یوازې د ټاکلو قواعدو له مخې ځوابونه ورکول، خو معاصر چېټبوټونه د لویو ژبنیو ماډلونو پر بنسټ روزل شوي دي. له همدې امله، دوی کولای شي اوږدې خبرې اترې ترسره کړي، علمي متنونه ولیکي، ژباړه وکړي، لنډیز برابر کړي، پروګرامونه ولیکي او د بېلابېلو موضوعاتو په اړه معلومات وړاندې کړي.
سره له دې، دغه سیستمونه تل بې‌خطا نه دي. کله ناکله داسې معلومات هم وړاندې کوي چې ظاهراً سم ښکاري، خو له واقعیت سره سمون نه لري. له همدې امله د علمي، حقوقي او طبي معلوماتو په برخه کې د انساني ارزونې اړتیا لا هم موجوده ده.

په تعلیم او څېړنه کې د محاسبوي ژبپوهنې تطبیق:

په تعلیمي برخه کې محاسبوي ژبپوهنه د هوښیارو ښوونیزو سیستمونو، اتومات ازموینو، د ژبې د زده کړې اپلیکېشنونو او برېښنایي کتابونو د جوړولو لپاره کارول کیږي. په علمي څېړنو کې د میلیونونو مقالو تحلیل، د مراجعو موندل، د متن لنډیز او د علمي معلوماتو طبقه‌بندي د دې علم له مهمو کارونو څخه دي.
د همدې پرمختګ له امله څېړونکي اوس کولای شي په لنډ وخت کې هغه معلومات ترلاسه کړي چې پخوا به یې موندلو ورځې او حتی میاشتې وخت نیوه.

د پښتو ژبې راتلونکې او اړتیاوې:

که څه هم د نړۍ په پرمختللو ژبو کې د محاسبوي ژبپوهنې ډېری تطبیقونه بریالي شوي، خو پښتو لا هم د ټیټو سرچینو ژبو له ډلې شمېرل کیږي. د دې وضعیت د بدلون لپاره د معیاري متني زېرمو جوړول، د لغوي او صرفي سرچینو پراخول، د غږیزو معلوماتو راټولول، د ژبنیو ټاګونو معیاري کول او د پوهنتونونو، څېړنیزو مرکزونو او ټکنالوژیکو بنسټونو ګډې هڅې اړینې دي.
که دا بنسټیز کارونه ترسره شي، نو پښتو به هم وکولای شي د نړۍ د پرمختللو ژبو په څېر د هوښیارو ژبنیو ټکنالوژیو بشپړه برخه وګرځي. دا به نه یوازې د ژبې د ساتنې او ودې سبب شي، بلکې د علم، اقتصاد، تعلیم او ډیجیټلي حکومتولۍ په پرمختګ کې به هم مهم رول ولوبوي.
محاسبوي ژبپوهنه او پښتو ژبه: علمي ننګونې، د ټیټو سرچینو ژبو ستونزې، د ژبنیو زېرمو جوړول او د مصنوعي ځیرکتیا لپاره د پښتو راتلونکې
پښتو ژبه د نړۍ له سترو سیمه‌ییزو ژبو څخه ده چې میلیونونه ویونکي لري او د افغانستان او پاکستان په پراخو سیمو کې د اړیکو، زده کړې، ادبیاتو او فرهنګ مهمه وسیله ګڼل کیږي. دغه ژبه د زرګونو کلونو فرهنګي او ادبي میراث لرونکې ده او د غني شعري، نثري او فولکلوري آثارو ښتنته ده. سره له دې، د ډیجیټلي ټکنالوژۍ او مصنوعي ځیرکتیا په نړۍ کې پښتو لا هم د هغو ژبو له ډلې شمېرل کیږي چې د محاسبوي سرچینو، معیاري معلوماتو او ژبنیو ټکنالوژیو له کمښت سره مخ دي. په همدې دلیل، د محاسبوي ژبپوهنې په برخه کې د پښتو علمي مطالعه او د هغې لپاره د معیاري محاسبوي سرچینو جوړول د وخت یوه اساسي اړتیا ده.
د نړۍ په معاصره علمي فضا کې د ژبې پرمختګ نور یوازې د ویونکو په شمېر نه ارزول کیږي، بلکې د هغې ډیجیټلي حضور، محاسبوي سرچینو، مصنوعي ځیرکتیا کې د کارونې وړتیا او د نړیوالو معلوماتي سیستمونو سره د یوځای کېدو ظرفیت هم مهم معیارونه ګڼل کیږي. که یوه ژبه په دغو برخو کې کمزورې وي، نو په تدریجي ډول به یې په نړیوالو علمي او ټکنالوژیکو تعاملاتو کې ونډه هم محدوده شي. له همدې امله، د پښتو لپاره د محاسبوي ژبپوهنې پراختیا یوازې یوه علمي اړتیا نه، بلکې یوه ستراتیژیکه او ملي اړتیا ده.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له راالیرلو مخکې په ځیر و لولئ

د کمو سرچینو لرونکي ژبې:(Low-Resource Languages)

په محاسبوي ژبپوهنه کې ژبې د موجودو ډیجیټلي سرچینو له مخې په دوو لویو ډلو وېشل کېږي: د لوړو سرچینو لرونکې ژبې (High-Resource Languages) او دکمو سرچینولرونکې ژبې (Low-Resource Languages).
د لوړو سرچینو ژبې، لکه انګلیسي، چینایي، فرانسوي، هسپانوي او جرمني، پراخ متني زېرمې، لوی لغتونه، غږیز معلومات، نحوي او ماناپوهنیز ډېټاسیټونه او پرمختللي ژبني ماډلونه لري. برعکس، د ټیټو سرچینو ژبې له دې امکاناتو څخه بې‌برخې وي. پښتو د محاسبوي ژبپوهنې له نظره د ټیټو سرچینو ژبو له ډلې شمېرل کېږي. د دې اصلي لاملونه د معیاري متني زېرمو کمښت، د ژبنیو معلوماتو محدودیت، د څېړنیزو پروژو نشتوالی او د متخصصو انساني سرچینو کموالی دی. همدا ستونزې د دې سبب شوې چې د پښتو لپاره د ماشین ژباړې، غږ پېژندنې، متن تولید او نورو هوښیارو سیستمونو کیفیت د پرمختللو ژبو په کچه نه وي.

د پښتو ژبې ژبنی ځانګړتیاوې:

د پښتو ژبې محاسبوي تحلیل له څو ځانګړو ستونزو سره مخ دی. لومړی، پښتو یوه صرفي پلوه بډایه (Morphologically Rich)ژبه ده. په دې ژبه کې د یوه فعل یا نوم لسګونه بڼې رامنځته کېدای شي، چې د جنس، شمېر، حالت، زمان او شخص له مخې بدلېږي. دا ځانګړنه د محاسبوي تحلیل پروسه پېچلې کوي.
دویم، پښتو نسبتاً ازاد نحوي ترتیب لري. که څه هم د جملې عادي جوړښت فاعل—مفعول—فعل (SOV) دی، خو د تینګار، سبک او ادبي اړتیا له مخې د جملې اجزا بدلېدای شي. دا انعطاف د نحوي تحلیل لپاره اضافي ستونزې رامنځته کوي.
درېیم، په پښتو کې د لهجو تنوع ډېره ده. د کندهاري، ننګرهاري، یوسفزي، وزیري، غلجي او نورو لهجو ترمنځ د تلفظ، لغتونو او ځینو ګرامري ځانګړتیاوو توپیرونه شته. که څه هم دا تنوع د ژبې فرهنګي بډاینه ده، خو د محاسبوي سیستمونو د روزنې لپاره د معیاري کولو اړتیا رامنځته کوي.

د متني زېرمو (Corpora) اهمیت:

د محاسبوي ژبپوهنې بنسټ د ژبنیو معلوماتو پر زېرمو ولاړ دی. هر څومره چې د یوې ژبې متني زېرمې پراخې، معیاري او متوازنې وي، په هماغه اندازه د هغې ژبې لپاره د مصنوعي ځیرکتیا سیستمونه دقیقې پایلې وړاندې کوي.
د پښتو لپاره د معیاري متني زېرمو جوړول باید د بېلابېلو ژبنیو سبکونو استازیتوب وکړي. په دې زېرمو کې باید علمي متنونه، ادبي آثار، ژورنالیستي لیکنې، حقوقي اسناد، تعلیمي کتابونه، رسمي مکاتبات، ټولنیزې رسنۍ او ورځنۍ محاورې شاملې وي. دا تنوع د دې سبب کېږي چې ژبني ماډلونه د ژبې بېلابېل سبکونه او کارونې زده کړي. متني زېرمې یوازې د متنونو ټولګه نه ده، بلکې باید د ژبنیو ځانګړتیاوو له مخې هم نښه‌ګذارۍ (Annotation) شي. د بېلګې په توګه، هره کلمه باید د خپل ګرامري ډول، صرفي ځانګړتیاوو، نحوي رول او ماناپوهنیزو اړیکو له مخې مشخصه شي. دا نښه‌ګذاري د ماشین زده کړې لپاره اساسي اهمیت لري.

د لغوي سرچینو جوړول:

د پښتو لپاره د محاسبوي ژبپوهنې یوه بله مهمه اړتیا د معیاري لغوي سرچینو رامنځته کول دي. په پرمختللو ژبو کې دWordNet ، اصطلاح‌نامو، مترادفونو، متضادو کلمو، معنايي شبکو او الکترونیکي قاموسونو پراخې زېرمې موجودې دي، چې د ژبنیو سیستمونو په روزنه کې مهم رول لري. د پښتو لپاره هم باید داسې علمي لغوي بانکونه جوړ شي چې پکې د کلمو ماناوې، مترادفونه، متضادونه، اصطلاحات، نحوي ځانګړتیاوې او د کارونې بېلګې ثبت شوې وي. دغه سرچینې نه یوازې د ماشین ژباړې کیفیت لوړوي، بلکې د پوښتنو ځوابوونکو سیستمونو، متن تولید، احساساتو تحلیل او هوښیارو قاموسونو د جوړولو لپاره هم بنسټیز ارزښت لري.

د غږیزو معلوماتو راټولول:

د غږ پېژندنې او مصنوعي وینا تولید لپاره د پراخو غږیزو معلوماتو شتون اړین دی. د پښتو لپاره باید د ښځو، نارینه‌وو، ماشومانو او د بېلابېلو سیمو د ویونکو غږونه په معیاري ډول ثبت شي. دا معلومات باید د عمر، جنس، لهجې او چاپېریال له مخې متوازن وي، څو روزل شوي ماډلونه په واقعي شرایطو کې هم لوړ دقت ولري. غږیز

معلومات باید د دقیقو متنونو سره یوځای ثبت شي، څکه د غږ او متن ترمنځ همغږي د غږ پېژندني او وینا تولید دواړو لپاره بنسټیز شرط دی.

د پښتو او مصنوعي ځیرکتیا اړیکه:

د لویو ژبنیو ماډلونو د پرمختګ سره، د پښتو لپاره هم نوي علمي او تخنیکي فرصتونه رامنځته شوي دي. که د پښتو لپاره معیاري ډیجیټلي معلومات برابر شي، نو مصنوعي ځیرکتیا به وکولای شي د پښتو متنونه تحلیل کړي، ژباړه ترسره کړي، علمي مقالې لنډې کړي، د زده‌کوونکو پوښتنو ته ځواب ووایي او د عامه خدماتو لپاره هوښیار سیستمونه وړاندې کړي. دا پرمختګ به یوازې د ټکنالوژۍ په برخه کې نه وي، بلکې د پښتو ژبې د ساتنې، معیاري کولو او نړیوال حضور په پیاوړتیا کې به هم مهم رول ولوبوي. ژبه چې ډیجیټلي حضور ونه لري، په تدریجي ډول له نړیوالو معلوماتي شبکو څخه څنګ ته کېږي.

د پوهنتونونو او څېړنیزو بنسټونو رول:

د پښتو د محاسبوي ژبپوهنې پرمختګ یوازې د ټکنالوژۍ د شرکتونو کار نه دی. پوهنتونونه، د ژبې اکاډمۍ، د معلوماتي ټکنالوژۍ مرکزونه او څېړنیز بنسټونه باید په ګډه کار وکړي. د کمپیوټر ساینس، ژبپوهنې او مصنوعي ځیرکتیا ګډ پروګرامونه، د محاسبوي ژبپوهنې ځانګړې څانګې، د معلوماتو د راټولولو ملي پروژې او د نړیوالو علمي مرکزونو سره همکاري د دې برخې د ودې لپاره مهم عوامل دي.
په ځانګړې توګه، د ماسترۍ او دوکتورا په کچه باید داسې څېړنې وهڅول شي چې د پښتو لپاره نحوي تحلیلګرونه، صرفي تحلیلګرونه، ژبني ماډلونه، غږ پېژندونکي او ماشین ژباړونکي رامنځته کړي. دا ډول علمي پانګونه به د راتلونکو کلونو لپاره د پښتو ژبې ډیجیټلي بنسټ پیاوړی کړي.

راتلونکې لیدلوری:

تمه کېږي چې په راتلونکو کلونو کې د مصنوعي ځیرکتیا پرمختګ به د ټیټو سرچینو ژبو لپاره هم نوې لارې پرانیزي. د څوژبنیو (Multilingual) او څوکاري (Multitask) ژبنیو ماډلونو پراختیا دا امکان برابروي چې هغه ژبې هم د هوښیارو سیستمونو برخه شي چې د معلوماتو محدودې سرچینې لري.
د پښتو لپاره دا فرصت هغه وخت په عملي بریا بدلېدای شي چې د ژبنیو معلوماتو راټولول، معیاري کول او شریکول په ملي کچه د یوې علمي ستراتیژۍ برخه وګرځي. که حکومت، پوهنتونونه، خصوصي سکټور او نړیوال علمي بنسټونه په ګډه همکاري وکړي، نو پښتو به وکولای شي د مصنوعي ځیرکتیا په عصر کې خپل مناسب ځای خپل کړي. په توله کې، د پښتو ژبې راتلونکې له محاسبوي ژبپوهنې سره ژوره اړیکه لري. هره پانګونه چې نن د ژبنیو سرچینو، علمي څېړنو او ډیجیټلي زیربناوو په جوړولو کې کېږي، په حقیقت کې د دې ژبې د علمي، فرهنګي او ټکنالوژیک راتلونکې لپاره پانګونه ده .
له همدې امله، د پښتو محاسبوي ژبپوهنه باید د ژبې د پرمختګ، د پوهې د تولید او د نړیوالو معلوماتي سیستمونو سره د اغېزمن ګډون لپاره یوه ملي لومړیتوب وګرځي.

د محاسبوي ژبپوهنې ننګونې، اخلاقي مسایل او راتلونکې:

د محاسبوي ژبپوهنې او مصنوعي ځیرکتیا بې‌ساري پرمختګونو د انسان او ټکنالوژۍ ترمنځ د اړیکو نوې نړۍ رامنځته کړې ده. نن ورځ هوښیار ژبني سیستمونه د ژباړې، تدریس، روغتیا، سوداګرۍ، رسنیو، قانون، امنیت او علمي څېړنو په برخو کې پراخه کارونه لري. سره له دې چې دغه پرمختګونه د بشري ژوند د اسانتیا، د معلوماتو د چټک لاسرسي او د ژبنیو خنډونو د کمولو سبب شوي، خو له بلې خوا یې ګڼې علمي، تخنیکي، حقوقي او اخلاقي ننګونې هم رامنځته کړې دي. له همدې امله، د محاسبوي ژبپوهنې راتلونکې یوازې د ټکنالوژۍ په پرمختګ پورې نه، بلکې د دغو ننګونو د سم مدیریت او د اخلاقي اصولو د رعایت پورې هم تړاو لري.

د ژبې ذاتي پېچلتیا:

د محاسبوي ژبپوهنې تر ټولوستره ننګونه د انسان د ژبې ذاتي پېچلتیا ده. طبیعي ژبه یوازې د لغتونو او ګرامري قواعدو مجموعه نه ده، بلکې د ټولنیز، فرهنګي، تاریخي او رواني عواملو تر اغېز لاندې یو ژوندی نظام دی. یوه

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

کلمه په بېلابېلو سیاقونو کې بېلابېلې ماناوې لرلای شي، یوه جمله کېدای شي هم مهرباني څرګنده کړي، هم طنز او هم غیر مستقیمه غوښتنه. د بېلګې په توګه، که یو څوک ووایي: «نن خو ډېر وختي راغلې!» دا جمله د ظاهري جوړښت له مخې ستاینه ښکاري، خو د سیاق له مخې کېدای شي طنز یا نیوکه وي. د دې ډول ماناوو دقیق درک لا هم د محاسبوي ژبپوهنې لپاره یوه مهمه علمي ستونزه ده.

د سیاق (Context) درک:

که څه هم لوی ژبني ماډلونه د سیاق په تحلیل کې د پام وړ پرمختګ کړی، خو بیا هم د انسان د پوهې، تجربې او تولنیزو اړیکو بشپړ درک نه لري. انسان د یوې جملې مانا د پخوانیو تجربو، فرهنګي ارزښتونو، ټولنیزو شرایطو او حتی د خبرو د ویلو له انداز څخه استنباطوي، خو ماشین تر ډېره د روزنیزو معلوماتو پر بنسټ احتمالي اړیکې زده کوي. همدا لامل دی چې ځینې وخت ژبني ماډلونه ظاهراً منطقي، خو له واقعي شرایطو سره ناسمه پایله وړاندې کوي. د محاسبوي ژبپوهنې راتلونکې څېړنې به تر ډېره د همدې ستونزې د حل لپاره متمرکزې وي، څو ماشین وکولای شي د ژبې ژوره مانا او ټولنیز سیاق په ښه ډول درک کړي.

د ژبنیو معلوماتو کمښت:

د نړۍ ډېری ژبې، په ځانګړې توګه د آسیا او افریقا ګڼې سیمه‌ییزې ژبې، د محاسبوي سرچینو له کمښت سره مخ دي. د لوړ کیفیت متني زېرمو، غږیزو معلوماتو، نحوي نښه‌ګذارۍ او ماناپوهنیزو سرچینو نشتوالی د دې سبب شوی چې د دغو ژبو لپاره ژبني ماډلونه د پرمختللو ژبو په کچه اغېزمن نه وي.
پښتو هم له همدې ستونزې سره مخ ده. د پراخو متني زېرمو، معیاري قاموسونو، موازي ژباړنیزو معلوماتو او غږیزو ډېټاسیټونو کمښت د دې لامل شوی چې د پښتو لپاره د مصنوعي ځیرکتیا ډېری سیستمونه لا هم د پرمختګ په لومړیو پړاوونو کې وي. د دې ستونزې د حل لپاره د ملي ژبنیو پروژو، ازادو ډیجیټلي زېرمو او نړیوالو علمي همکاریو پراختیا اړینه ده.

د ژبنیو ماډلونو تعصب:(Bias)

د محاسبوي ژبپوهنې یوه مهمه اخلاقي ستونزه د ژبنیو ماډلونو تعصب دی. لوی ژبني ماډلونه د هغو معلوماتو له مخې روزل کېږي چې په روزنیزو زېرمو کې موجود وي. که دا معلومات یو اړخیز، ناسم یا له تعصب څخه ډک وي، نو ماډل هم ورته تمایلات خپلوي.
دغه تعصب کېدای شي د جنسیت، ژبې، قوم، کلتور، ټولنیز طبقې یا جغرافیایي سیمو په اړه وي. د بېلګې په توګه، که یوه ژبه د روزنیزو معلوماتو په زېرمو کې لږه استازولي ولري، نو د هغې ژبې لپاره د ژبني ماډل کیفیت هم ټیټېږي. له همدې امله، معاصرې څېړنې هڅه کوي چې متوازني او متنوع ژبنۍ زېرمې رامنځته کړي، څو د ژبنیو ماډلونو تعصب کم شي او د بېلابېلو ژبو او کلتورونو عادلانه استازیتوب تضمین شي.

د معلوماتو محرمیت او امنیت:

د مصنوعي ځیرکتیا د پراخ استعمال له امله د شخصي معلوماتو خوندیتوب یوه جدي اندېښنه ده. د ژبنیو سیستمونو د روزنې لپاره کله ناکله د کاروونکو متنونه، غږونه او نور معلومات کارول کېږي. که د دغو معلوماتو د ساتنې لپاره مناسب حقوقي او تخنیکي چوکاټ موجود نه وي، نو د محرمیت د نقض خطر رامنځته کېږي.
له همدې امله، د محاسبوي ژبپوهنې د راتلونکي پرمختګ لپاره اړینه ده چې د معلوماتو راټولول، ساتنه او کارونه د شفافو اخلاقي او حقوقي اصولو له مخې ترسره شي، څو د افرادو شخصي حقونه خوندي پاتې شي.

ناسم معلومات او د حقیقت ارزونې ستونزه:

لوی ژبني ماډلونه ځینې وخت داسې معلومات تولیدوي چې له ژبني پلوه سم او روان وي، خو له علمي یا واقعي پلوه دقیق نه وي. په مصنوعي ځیرکتیا کې دې حالت ته د Hallucination اصطلاح کارول کېږي. دا پدیده په ځانګړې توګه په علمي، حقوقي او طبي برخو کې جدي ستونزې رامنځته کولای شي. له همدې امله، د ژبنیو ماډلونو ځوابونه باید د معتبر علمي سرچینو او متخصصانو له خوا وارزول شي. مصنوعي ځیرکتیا د انساني متخصص ځای نه شي نیولای، بلکې د هغه د کار د اسانتیا او چټکتیا لپاره یوه مرستندویه وسیله ده.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولـئ

اخلاقي مسؤلیت:

د محاسبوي ژبپوهنې پرمختګ د علمي مسؤلیتونو نوی باب هم پرانیستی دی. څېړونکي، پروګرام جوړونکي، پوهنتونونه او ټکنالوژیک شرکتونه مسؤلیت لري چې داسې سیستمونه جوړ کړي چې د انسان کرامت، فرهنګي تنوع، ژبنۍ عدالت او ټولنیز مسؤلیت اصول پکې رعایت شوي وي. د مصنوعي ځیرکتیا هره نوې ټکنالوژي باید د بشري ارزښتونو د پیاوړتیا لپاره وکارول شي، نه د تبعیض، ناسمو معلوماتو یا د ټولنیزو ستونزو د زیاتولو لپاره. له همدې امله، د «مسؤلې مصنوعي ځیرکتیا (Responsible AI)» مفهوم په وروستیو کلونو کې د نړیوالو علمي بحثونو مهمه موضوع ګرځېدلې ده.

د محاسبوي ژبپوهنې راتلونکی:

د محاسبوي ژبپوهنې راتلونکی له بېسارو فرصتونو ډک ښکاري. تمه کېږي چې په راتلونکو کلونو کې به ژبني ماډلونه د انسان د ژبې په درک، استدلال، استنباط او څوژبني تعامل کې لا ډېر پرمختګ وکړي. د غږ، متن، انځور او ویډیو ګډ تحلیل به د څووسیلېز (Multimodal) مصنوعي ځیرکتیا نوی پړاو رامنځته کړي.
په تعلیمي برخه کې به هوښیار ښوونیز مرستیالان د هر زدهکوونکي د اړتیا له مخې ځانګړی تعلیمي محتوا چمتو کړي. په روغتیا کې به د ناروغانو راپورونه په اتومات ډول تحلیل شي. په حقوقي چارو کې به د اسنادو ارزونه او د قضیو لومړنۍ شننه په چټکۍ سره ترسره کېږي. په علمي څېړنو کې به مصنوعي ځیرکتیا د معلوماتو د تحلیل، د څېړنیزو تشو د پېژندنې او د نوو فرضیو د وړاندیز په برخه کې مهم رول ولوبوي. د ټیټو سرچینو ژبو لپاره هم راتلونکی هیلهبښونکی دی. د څوژبنیو ژبنیو ماډلونو پرمختګ دا زمینه برابروي چې پښتو، ازبکي، بلوچي، ترکمني او نورې سیمهییزې ژبې هم د نړیوالو ژبنیو ټکنالوژیو برخه شي، خو دا پرمختګ یوازې هغه وخت عملي کېږي چې د دغو ژبو لپاره معیاري متني او غږیزې سرچینې رامنځته شي.

د افغانستان لپاره ستراتیژیک اهمیت:

د افغانستان لپاره محاسبوي ژبپوهنه یوازې یوه علمي څانګه نه، بلکې د علمي پرمختګ، ډیجیټلي خپلواکۍ او ژبني عدالت وسیله ده. د پښتو او دري لپاره د ملي ژبنیو زېرمو جوړول، د مصنوعي ځیرکتیا د څېړنیزو مرکزونو رامنځته کول، د پوهنتونونو ترمنځ همکاري، او د ژبپوهانو او کمپیوټرپوهانو ګډې څېړنې کولای شي د هېواد د علمي پرمختګ لپاره نوې لارې پرانیزي. په ځانګړې توګه، د محاسبوي ژبپوهنې تدریس باید د ژبپوهنې، کمپیوټر ساینس او مصنوعي ځیرکتیا د تحصیلي پروګرامونو یوه مهمه برخه وګرځي. د ماسټرۍ او دوکتورا په کچه څېړنې، د ژبنیو معلوماتو راټولول، د خلاصو (Open) ژبنیو سرچینو جوړول او د نړیوالو علمي مرکزونو سره همکاري به د افغانستان د ژبنیو ټکنالوژیو راتلونکی پیاوړی کړي.

پایيزه ارزونه:

د محاسبوي ژبپوهنې پرمختګ څرګندوي چې ژبه نور یوازې د ادبیاتو او دودیزې ژبپوهنې موضوع نه ده، بلکې د مصنوعي ځیرکتیا، معلوماتي ټکنالوژۍ او ډیجیټلي اقتصاد اساسي برخه ګرځېدلې ده. سره له دې چې د ژبې ذاتي پېچلتیا، د معلوماتو کمښت، تعصب، محرمیت او د حقیقت د ارزونې ستونزې لا هم موجودې دي، خو د نړیوالو علمي هڅو دوام دا هیله پیاوړې کوي چې راتلونکي ژبني سیستمونه به لا دقیق، عادلانه، شفاف او د انسان د اړتیاوو سره ډېر همغږي وي. د پښتو ژبې لپاره هم دا یوه تاریخي موقع ده. که نن د ژبنیو سرچینو، علمي څېړنو او محاسبوي زیربناوو په جوړولو کې پانګونه وشي، نو راتلونکی نسل به وکولای شي چې خپله ژبه د مصنوعي ځیرکتیا، نړیوالو علمي شبکو او ډیجیټلي نړۍ په ټولو برخو کې په بشپړ ځواک وکاروي. همدې موخې ته رسېدل د محاسبوي ژبپوهنې د راتلونکي تر ټولو ستره لاسته راوړنه ګڼل کېږي.

وړاندیزونه:(Recommendations)

د دې څېړنې د موندنو او تحلیلونو پر بنسټ لاندې علمي او عملي وړاندیزونه وړاندې کېږي:

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بڼې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

۱ .د افغانستان په پوهنتونونو کې باید محاسبوي ژبپوهنه (Computational Linguistics) د ژبپوهنې، کمپیوټر ساینس او مصنوعي ځیرکتیا د نصاب یوه اساسي برخه وګرځول شي، څو محصلین د ژبې او ټکنالوژۍ د اړیکو په اړه معاصرې علمي پوهې ته لاسرسی ومومي.
۲ .د پښتو ژبې لپاره باید پراخ او معیاري متني زېرمې (Corpora) جوړې شي، چې ادبي، علمي، رسنيز، حقوقي، طبي او ورځني متنونه پکې شامل وي، څکه د ژبنیو ماډلونو د کیفیت بنسټ د معیاري معلوماتو شتون دی.
۳ .د پښتو ژبې لپاره باید د صرفي تحلیلګر(Morphological Analyzer) ، نحوي تحلیلګر(Parser) ، برېښنايي قاموسونو، WordNetاو نورو ژبنیو سرچینو جوړولو ته لومړیتوب ورکړل شي.
۴ .د لوړو زده کړو وزارت، د اطلاعاتو او کلتور وزارت، د افغانستان علومو اکاډمۍ او اړوندو علمي بنسټونو ترمنځ باید ګډې علمي پروژې رامنځته شي، څو د پښتو محاسبوي سرچینې په ملي کچه تولید او معیاري شي.
۵ .د مصنوعي ځیرکتیا او محاسبوي ژبپوهنې په برخه کې باید د ماسټرۍ او دوکتورا څېړنې وهڅول شي، په ځانګړې توګه د پښتو د ټیټو سرچینو ژبې (Low-Resource Language) د ستونزو د حل لپاره.
۶ .د پښتو ژبې د غږ پېژندنې(Speech Recognition) ، غږ تولید (Text-to-Speech) او ماشین ژباړې لپاره باید معیاري غږیز او متني معلومات راټول او نښه‌ګذاري شي.
۷ .د پښتو ژبې لپاره د لویو ژبنیو ماډلونو (Large Language Models) د روزنې په موخه باید د خلاصو (Open-source)ژبنیو معلوماتو ملي زېرمه رامنځته شي.
۸ .د پوهنتونونو د ژبپوهنې، کمپیوټر ساینس او معلوماتي ټکنالوژۍ پوهنځیو ترمنځ باید بین‌رشتوي (Interdisciplinary)همکارۍ پراخه شي.
۹ .د محاسبوي ژبپوهنې په ټولو پروژو کې باید د اخلاقي اصولو، د معلوماتو د محرمیت، ژبني عدالت او د مصنوعي ځیرکتیا د مسؤلانه کارونې معیارونه مراعات شي.
۱۰ .د پښتو ژبې د نړیوال حضور د پیاوړتیا لپاره باید د نړیوالو علمي مرکزونو، پوهنتونونو او څېړنیزو بنسټونو سره دوامداره همکارۍ رامنځته شي.

پایله:

د دې څېړنې تحلیل دا په ډاګه کوي، چې محاسبوي ژبپوهنه د معاصرې ژبپوهنې، کمپیوټر ساینس او مصنوعي ځیرکتیا له مهمو علمي څانګو څخه ده، چې د انسان او ماشین ترمنځ د طبیعي اړیکو د رامنځته کولو لپاره بنسټیز رول لري. د قواعدو پر بنسټ له لومړنیو سیستمونو څخه نیولې د ژورو عصبي شبکو او لویو ژبنیو ماډلونو تر دورې پورې، دغه علم د پام وړ پرمختګ کړی او نن ورځ د ژباړې، غږ پېژندنې، متن تولید، معلوماتو راایستلو، احساساتو تحلیل او هوښیارو مرستیالانو په څېر ګنو برخو کې عملي کارول کېږي.
څېړنه څرګندوي چې پښتو ژبه، سره له دې چې د ویونکو له پلوه یوه لویه ژبه ده، د محاسبوي ژبپوهنې له نظره د ټیټو سرچینو ژبو په ډله کې راځي. د متني او غږیزو زېرمو کمښت، د معیاري ژبنیو سرچینو نشتوالی او د بین‌رشتوي څېړنو محدودیتونه د دې سبب شوي چې د پښتو لپاره د مصنوعي ځیرکتیا پرمختللي سیستمونه لا هم د ودې په پړاو کې وي. سره له دې، د څوژبنیو ماډلونو، ژورو زده کړو او تولیدي مصنوعي ځیرکتیا وروستي پرمختګونه د پښتو ژبې لپاره نوي فرصتونه رامنځته کړي دي. که علمي بنسټونه، پوهنتونونه، دولتي اداري او خصوصي سکټور په ګډه د ژبنیو سرچینو په جوړولو او څېړنو کې پانګونه وکړي، نو پښتو به وکولای شي چې د نړیوالو ژبنیو ټکنالوژیو په ډګر کې خپل مناسب ځای خپل کړي. محاسبوي ژبپوهنه نه یوازې د ژبې د تحلیل یوه علمي وسیله ده، بلکې د راتلونکې ډیجیټلې ټولنې، مصنوعي ځیرکتیا او د پوهې د اقتصاد د جوړولو اساسي بنسټ هم ګڼل کېږي.

References

1. Allen, J. (1995). Natural Language Understanding (2nd ed.). Benjamin/Cummings.
2. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
3. Chomsky, N. (1957). Syntactic Structures. Mouton.
4. Chomsky, N. (1965). Aspects of the Theory of Syntax. MIT Press.
5. Chomsky, N. (1981). Lectures on Government and Binding. Foris Publications.
6. Clark, A., Fox, C., & Lappin, S. (2010). The Handbook of Computational Linguistics and Natural Language Processing. Wiley-Blackwell.
7. Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. Morgan & Claypool.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکنې د لیکنیزې بنې پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له رالیږلو مخکې په ځیر و لولئ

8. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
9. Jurafsky, D., & Martin, J. H. (2025). Speech and Language Processing (3rd ed., draft). Pearson.
10. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
11. Manning, C. D. (2022). Human Language Understanding & AI. Stanford University.
12. Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
13. Saussure, F. de. (1983). Course in General Linguistics. Duckworth.
14. Shannon, C. E. (1948). "A Mathematical Theory of Communication." Bell System Technical Journal, 27, 379–423.
15. Turing, A. M. (1950). "Computing Machinery and Intelligence." Mind, 59(236), 433–460.
16. Vaswani, A., et al. (2017). "Attention Is All You Need." Advances in Neural Information Processing Systems (NeurIPS).
17. Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). "A Neural Probabilistic Language Model." Journal of Machine Learning Research, 3, 1137–1155.
18. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL-HLT.
19. Brown, T. B., et al. (2020). "Language Models are Few-Shot Learners." NeurIPS.
20. OpenAI. (2023). GPT-4 Technical Report.
21. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). "Efficient Estimation of Word Representations in Vector Space."
22. Goldberg, Y. (2016). "A Primer on Neural Network Models for Natural Language Processing." Journal of Artificial Intelligence Research.
23. Liu, Y., et al. (2019). "RoBERTa: A Robustly Optimized BERT Pretraining Approach."
24. Lewis, M., et al. (2020). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation."
25. Raffel, C., et al. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer."
26. Radford, A., et al. (2019). "Language Models are Unsupervised Multitask Learners."
27. Koehn, P. (2010). Statistical Machine Translation. Cambridge University Press.
28. Sproat, R. (1992). Morphology and Computation. MIT Press.
29. Nivre, J. (2005). Dependency Grammar and Dependency Parsing.
30. Winograd, T. (1972). Understanding Natural Language. Academic Press.
31. Jurafsky, D. (2023). Computational Linguistics Lecture Notes. Stanford University.
32. Manning, C. D. (2023). CS224N: Natural Language Processing with Deep Learning. Stanford University.
33. Eisenstein, J. (2019). Introduction to Natural Language Processing. MIT Press.
34. Jurafsky, D. (2000). "Speech & Language Processing." Pearson.
35. Cambria, E., & White, B. (2014). "Jumping NLP Curves: A Review of Natural Language Processing Research."
36. Hirschberg, J., & Manning, C. D. (2015). "Advances in Natural Language Processing." Science, 349(6245), 261–266.
37. Jurafsky, D. (2024). Large Language Models and NLP. Stanford University.
38. Goodfellow, I. (2023). Recent Advances in Deep Learning.
39. Oxford Handbook of Computational Linguistics. (2005). Oxford University Press.
40. Mitkov, R. (Ed.). (2003). The Oxford Handbook of Computational Linguistics. Oxford University Press.

افغان جرمن آنلاین په درنښت تاسو همکارۍ ته رابولي. په دغه پته له موږ سره اړیکه ټینګه کړئ maqalat@afghan-german.de
یادونه: دلیکني د لیکنیزې بني پازوالي د لیکوال په غاړه ده ، هیله من یو خپله لیکنه له راالیرلو مخکي په ځیر و لولئ