International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly UGC Care Listed
Vol. 6 No. 41 November - December - 2025

తెలుగు భాషా సాంకేతికత: ప్రస్తుత స్థితి, సవాళ్లు మరియు అభివృద్ధి అవకాశాలు

Samudrala Samatha
Lecturer in Telugu,
Government Degree College Godavarikhani, District: Peddapelli, State: Telangana
Corresponding Author – Samudrala Samatha
DOI - 10.5281/zenodo.17726395

సారాంశం:

డిజిటల్ యుగంలో భాషా సాంకేతికతకు అత్యంత ప్రాధాన్యం ఏర్పడింది. ప్రత్యేకంగా తెలుగు వంటి ద్రావిడ భాషలు తమ విస్తృత వ్యాకరణ వ్యవస్థ, లిపి సంక్లిష్టత, ధ్వన్యాత్మక లక్షణాల వల్ల ప్రత్యేక పరిశోధనకు అర్థమైనవి. తెలుగు భాషా సాంకేతికతలో యంత్రానువాదం, వాక్య గుర్తింపు, వాక్య నిర్మాణ విశ్లేషణ, అక్షరదోష నిర్ధారణ, నిఘంటువుల నిర్మాణం, సహజ భాషా ప్రాసెసింగ్ (NLP) మోడళ్ల అభివృద్ధి వంటి రంగాలు వేగంగా అభివృద్ధి చెందుతున్నాయి. ఈ పత్రం ద్వారా తెలుగు భాషా సాంకేతికత యొక్క ప్రస్తుత స్థితి, సవాళ్లు మరియు అభివృద్ధి మార్గాలను విశ్లేషించడం లక్ష్యం.

కీవర్డ్లు: తెలుగు భాషా సాంకేతికత, యంత్రానువాదం, NLP, వాక్య విశ్లేషణ, ధ్వనిశాస్త్రం, లిపిశాస్త్రం.

పరిచయం:

ఇరవై ఒకటవ శతాబ్దంలో ప్రపంచ జ్ఞాన వ్యవస్థ మొత్తం డిజిటల్ రంగంలోకి వేగంగా చేరుతున్న నేపథ్యంలో, భాషలకు సంబంధించిన సాంకేతికత (Language Technology) అనేది అత్యంత కీలకమైన పరిశోధనా విభాగంగా ఎదిగింది. ఈ నేపథ్యంలో భారతదేశంలోని ప్రధాన ద్రావిడ భాషలలో ఒకటైన **తెలుగు**, తన ప్రత్యేక లిపి నిర్మాణం, పదవిన్యాసం, సంధి–సమాసాల సమృద్ధి, ధ్వన్యాత్మక లక్షణాలతో అనేక భాషా సాంకేతిక అన్వయాలకు విస్తృతమైన అవకాశాలను అందిస్తుంది.

తెలుగు మాట్లాడేవారి సంఖ్య ప్రపంచవ్యాప్తంగా **8 కోట్లకుపైగా** ఉండటం, ఇంటర్నెట్ వినియోగం పెరగడం, ఆన్లైన్ విద్య, డిజిటల్ ప్రభుత్వ సేవలు, సోషల్‌మీడియా విస్తరణ—all these—తెలుగు భాషా సాంకేతికతను అత్యవసరంగా చేయిస్తున్నాయి. భాషను యంత్రం అర్థం చేసుకోవడం, విశ్లేషించడం, ఉత్పత్తి చేయడం వంటి పనులకు సహజ భాషా ప్రాసెసింగ్ (NLP), యంత్రానువాదం (Machine

Translation), మాట గుర్తింపు (Speech Recognition), పాఠ్య–వాక్య ఉత్పత్తి (Text and Speech Generation), OCR (Optical Character Recognition) వంటి సాంకేతికతలు అవసరం.

తెలుగు భాష విషయంలో ఈ సాంకేతికతలను అభివృద్ధి చేయడం అంత సులభం కాదు. కారణం—తెలుగు పదనిర్మాణపు ప్రత్యేకతలు, విస్తారమైన రూపాంతరాలు, శబ్ద నిర్మాణంలోని నుడికట్టు, లిపి సంక్లిష్టత, వ్రాత–ఉచ్చారణ మధ్య తేడాలు, సంధులు, సమాసాలు, విపులమైన వ్యాకరణ వ్యవస్థ. ఇవన్నీ కంప్యూటర్లకు సవాళ్ళను సృష్టిస్తున్నప్పటికీ, అదే సమయంలో పరిశోధనకు కొత్త దారులను తెరవుతున్నాయి.

ఈ నేపథ్యంలో **తెలుగు భాషా సాంకేతికత**పై పరిశోధన అనేది కింది అంశాల గురించి లోతుగా అర్థం చేసుకోవడం అవసరం:

- తెలుగు భాషను డిజిటల్ దృష్టిలో ఎలా ప్రాతినిధ్యం ఇవ్వాలి?
- యంత్రానువాదం, వాక్య గుర్తింపు వంటి రంగాలలో ప్రస్తుతం ఉన్న పురోగతి ఎంత?
- NLP మోడళ్ల పనితీరు మెరుగుపర్చడానికి అవసరమైన డేటా వనరులు ఏమిటి?
- తెలుగు భాషకు ప్రత్యేకంగా సరిపోయే అల్గోరిథంలను ఎలా రూపొం

టించాలి?

- భవిష్యత్తులో AI ఆధారిత పెద్ద భాషా నమూనాలు (LLMs) తెలుగు కోసం ఎలా అభివృద్ధి చేయబడవచ్చు?

ప్రస్తుతం IIIT-హైదరాబాద్, హైదరాబాద్ సెంట్రల్ యూనివర్సిటీ, DRDO, TDIL (Technology Development for Indian Languages), Microsoft, Google వంటి అనేక సంస్థలు తెలుగు భాషా సాంకేతిక పరిశోధనలో పాలుపంచుకుంటున్నాయి.

అయినప్పటికీ, భాషా డేటా కొరత, ప్రామాణిక నిఘంటువుల లోపం, అనోటేషన్ ఖర్చులు, లిపి–శబ్ద విభిన్నత వంటి సమస్యలు ఇంకా కొనసాగుతూనే ఉన్నాయి.

ఈ పరిశోధనా పత్రం ద్వారా **తెలుగు భాషా సాంకేతికత యొక్క ప్రస్తుత స్థితి, ఎదురయ్యే సవాళ్లు, వినియోగ రంగాలు, భవిష్యత్ అవకాశాలు, మరియు అభివృద్ధికి దోహదపడే మార్గాలను** సమగ్రంగా విశ్లేషించడం లక్ష్యం. ప్రత్యేకించి, తెలుగు NLP ప్రయోజనాలు, యంత్రానువాద వ్యవస్థల సామర్థ్యం, వాక్య నిర్మాణ విశ్లేషణ, వాక్య గుర్తింపు, OCR అభివృద్ధి వంటి అంశాలను పరిశీలించి, భవిష్యత్ పరిశోధనకు సిఫార్సులు అందించడం కూడా ఈ పత్రం యొక్క ప్రధాన ఉద్దేశ్యం.

సాహిత్య సమీక్ష:

తెలుగు భాషా సాంకేతికతపై గత రెండు దశాబ్దాలుగా అనేక పరిశోధనలు చోటు చేసుకున్నాయి. ముఖ్యంగా:

1. యంత్రానువాదం (Machine Translation):

- "శ్రీయంత్ర" మరియు "అనువాదకుడు" వంటి ప్రాజెక్టులు తెలుగు-హిందీ, తెలుగు-ఇంగ్లీష్ అనువాదంలో కీలక పాత్ర పోషించాయి.
- గూగుల్ ట్రాన్స్లేట్ తెలుగు భాషను చేర్చడం ద్వారా విస్తృత డేటా సేకరణ వేగవంతమైంది.

2. సహజ భాషా ప్రాసెసింగ్:

తెలుగు NLP లో క్రిందివి ప్రధాన పరిశోధనా రంగాలు:

- వాక్యభాగ విశ్లేషణ (POS Tagging)
- నేమ్డ్ ఎంటిటీ రికగ్నిషన్
- సెంటిమెంట్ అనాలిసిస్
- పెద్ద భాషా నమూనాలు (LLMs)

3. వాక్య గుర్తింపు (Speech Recognition):

- గూగుల్, మైక్రోసాఫ్ట్ సంస్థలు తెలుగు వాయిస్ ఇన్పుట్‌ను అభివృద్ధి చేశాయి.
- ధ్వన్యాత్మక విశ్లేషణకు సంబంధించిన అనేక అధ్యయనాలు హైదరాబాద్ సెంట్రల్ యూనివర్సిటీ, IIIT-హైదరాబాద్ వంటి సంస్థల్లో జరిగాయి.

4. అక్షరదోష నిర్ధారణ & లిపి ప్రాసెసింగ్

తెలుగు లిపి సంక్లిష్టత వల్ల OCR (Optical Character Recognition) లో అనేక సవాళ్లు ఎదురయ్యాయి. అయినప్పటికీ, కొత్త డీప్ లెర్నింగ్ మోడల్‌లతో ప్రస్తుతం ఖచ్చితత్వం గణనీయంగా మెరుగుపడింది.

పరిశోధనా విధానం:

ఈ పత్రం పూర్తిగా వివిధ పరిశోధనా పత్రాలు, సాంకేతిక నివేదికలు, ఆన్లైన్ వనరులను ఆధారంగా చేసుకొని చేసిన విశ్లేషణాత్మక అధ్యయనం. ముఖ్యంగా:

- సెకండరీ డేటా అనాలిసిస్
- కేస్ స్టడీ పరిశీలన
- ప్రస్తుతం వినియోగంలో ఉన్న AI సాధనాల సమీక్ష
- డేటాసెట్‌ల ఆధారిత పోలిక

డేటా విశ్లేషణ మరియు పరిశీలనలు:

1. సాంకేతిక పురోగతి:

తెలుగు భాషలో NLP సంబంధిత ప్రాజెక్టులు వేగంగా అభివృద్ధి చెందుతున్నాయి, ముఖ్యంగా:

- Transformer-based models (BERT, GPT ఆధారిత మోడళ్లు)
- Speech-to-text మోడళ్లు
- OCR సిస్టమ్‌లు
- ఎడ్యుకేషనల్ టూల్స్ (వ్యాకరణ పరిశీలన, వాక్య సరిచూపు)

2. ఎదురవుతున్న సవాళ్లు:

- భారీ స్థాయి అనోటేషన్ డేటా కొరత
- లిపి క్లిష్టత వల్ల OCR ఖచ్చితత్వం తగ్గడం
- సంధి, సమాసాల వల్ల పద విభజనలో సమస్యలు
- భాషా తేడాల వల్ల యంత్రానువాదంలో వాక్య నిర్మాణ సమస్యలు

చర్చ:

తెలుగు భాషా సాంకేతికత యొక్క అభివృద్ధి భారతీయ భాషా పరిశోధనలో ఒక కీలక మలుపు. ఈ రంగం అభివృద్ధి కేవలం సాంకేతిక పురోగతికి మాత్రమే పరిమితం కాకుండా, భాషా సంరక్షణ, డిజిటల్ అక్షరాస్యత, విద్యా విస్తరణ, సామాజిక చేరువ, పరిపాలనా సేవల అందుబాటుకు కూడా ప్రాముఖ్యమైన ప్రభావాన్ని చూపుతోంది. ఈ చర్చలో తెలుగు భాషా సాంకేతికతలోని ప్రధాన అంశాలు, పురోగతి, సవాళ్లు, అవకాశాలు మరియు భవిష్యత్ దిశలను విశ్లేషిస్తాము.

1. తెలుగు భాష యొక్క నిర్మాణ స్వరూపం—సాంకేతికతకు సవాళ్లు మరియు అవకాశాలు:

తెలుగు భాషలోని ధ్వన్యాత్మక స్పష్టత, అక్షరాల శాస్త్రీయ నిర్మాణం, పదరూపాల సమృద్ధి NLP మోడళ్లకు ప్రత్యేక అవకాశాలు కల్పిస్తాయి. అయితే సంధులు, విభక్తులు, సమాసాలు, క్రియారూపాల అధికత యంత్ర విశ్లేషణకు కష్టతలను తెస్తాయి. ఉదాహరణకు, "చేస్తున్నాను", "చేశాను", "చేయబోతున్నాను" వంటి క్రియారూపాలు యాంత్రిక విభజనలో సమస్యలు సృష్టిస్తాయి. ఇదే సమయంలో, వీటి స్పష్టమైన నియమబద్ధత డేటా augmentation మరియు morphological tagging లో ప్రయోజనకరంగా మారుతుంది.

2. యంత్రానువాద వ్యవస్థల సామర్థ్యం—ప్రస్తుత స్థితి:

తెలుగు–ఇంగ్లీష్ మరియు తెలుగు–హిందీ యంత్రానువాదాలలో గత దశాబ్దంలో డీప్ లెర్నింగ్ ప్రయోగాల వల్ల accuracy గణనీయంగా మెరుగుపడింది. Transformer ఆధారిత మోడళ్ల ప్రవేశం అనువాద నాణ్యతను ముందేకంటే మెరుగుపర్చింది. అయితే పలు సవాళ్లు ఇంకా కొనసాగుతున్నాయి:

- సామెతలు, ప్రయోగాలు, వాడుక భాష అనువాదంలో అస్పష్టత
- పొడవాటి వాక్యాల sentence alignment లో సమస్యలు
- పదార్థ వైవిధ్యం (polysemy) వల్ల అనువాద లోపాలు
- సంధి విభజనలో యంత్రాల దోషాలు

ఈ ప్రాంతాలు యంత్రానువాద పరిశోధనకు భవిష్యత్ మార్గాలను సూచిస్తున్నాయి.

3. వాక్య గుర్తింపు (Speech Recognition)—ఎదురైన ఇబ్బందులు:

తెలుగు ఒక ధ్వన్యాత్మక భాష అయినప్పటికీ, వాక్య గుర్తింపు వ్యవస్థలకు కొన్ని ప్రత్యేక సవాళ్లు ఉన్నాయి:

- ప్రాంతీయ ఉచ్చారణలు (ఉదా: కోస్తా, రాయలసీమ, తెలంగాణ)
- వేగంగా మాట్లాడే సమయంలో అక్షరాల కలయిక
- నేపథ్య శబ్దం ప్రభావం
- బాలల–మహిళల–వృద్ధుల స్వర భేదాలు

తాజా AI మోడళ్లతో ఈ సమస్యలు తగ్గుతున్నప్పటికీ, పెద్ద పరిమాణంలో శబ్ద డేటా అవసరం ఇంకా ఉంది.

4. తెలుగు OCR—పురోగతి మరియు పరిమితులు:

తెలుగు లిపి సమగ్రత, అక్షరాల కలయికలు, గుణింతాలు—OCR సిస్టమ్ లకు ప్రధాన సవాళ్లు. గతంలో accuracy తక్కువగా ఉన్నా ప్రస్తుతం CNN, LSTM, Vision Transformers ఉపయోగించడం వల్ల OCR ఖచ్చితత్వం పెరిగింది. అయినా OCR వ్యవస్థలు కిందివారితో ఇంకా ఇబ్బంది పడుతున్నాయి:

- పురాతన గ్రంథాల స్కాన్ చిత్రాలు
- చేతిరాత పత్రాలు
- ప్రత్యేక కళా అక్షరాలు
- ముద్రణా వైవిధ్యం

ఇవి భవిష్యత్ పరిశోధనలో ముఖ్య ప్రాంతాలుగా నిలుస్తాయి.

5. NLP మోడళ్ల అభివృద్ధి—LLMs మరియు తెలుగు భాష

GPT, BERT వంటి పెద్ద భాషా నమూనాలు (LLMs) అనేక భాషా పనులను ఆటోమేటిక్‌గా చేయగలగడం వాస్తవంగా భాషా సాంకేతికతలో విప్లవాత్మక మార్పులను తెచ్చాయి. తెలుగు కోసం LLMలను అభివృద్ధి చేయడంలో పరిమితులు:

- విశ్వసనీయంగా అనోటేట్ చేసిన పెద్ద కోర్పస్ లేకపోవడం
- డేటా వైవిధ్యం తక్కువగా ఉండటం
- డొమైన్-స్పెసిఫిక్ టెక్స్ట్ కొరత అయితే భవిష్యత్తులో

ప్రభుత్వం–విశ్వవిద్యాలయాలు–ప్రైవేట్ రంగం కలసి పనిచేస్తే తెలుగు LLMలు ప్రపంచ ప్రమాణాలను చేరగలవు.

6. సామాజిక, శిక్షణ మరియు పరిపాలనా రంగాల్లో తెలుగు భాషా సాంకేతికత ప్రయోజనాలు:

తెలుగు భాషా సాంకేతికత విస్తృతంగా సామాజిక జీవితంలో మార్పులను తీసుకువస్తోంది:

- e-governance సేవల్లో తెలుగు ఇన్‌పుట్
- విద్యా రంగంలో AI ఆధారిత అభ్యాస పరికరాలు
- ప్రత్యేక అవసరాలున్నవారికి వాయిస్–టెక్స్ట్ సహాయ సాఫ్ట్‌వేర్
- కంటెంట్ సృష్టిలో ఆటోమేషన్
- సోషల్ మీడియా మానిటరింగ్ మరియు భావ వ్యక్తీకరణ విశ్లేషణ

ఇవి భాషా సాంకేతికత సామాజిక ప్రభావాన్ని స్పష్టంగా సూచిస్తున్నాయి.

7. భవిష్యత్ అవకాశాలు:

భవిష్యత్తులో తెలుగు భాషా సాంకేతిక రంగంలో భారీగా అభివృద్ధి అవకాశాలు ఉన్నాయి:

- సంపూర్ణ conversational AI in Telugu
- డొమైన్ స్పెసిఫిక్ chatbots (వైద్యం, వ్యవసాయం, విద్య, చట్టం)
- అధునాతన యంత్రానువాదం
- చేతిరాత recognition
- నిఘంటువుల మరియు భాషా డేటాబేస్‌ల నిర్మాణం
- భాషా బోధనలో AI-based tutor systems

సారాంశాత్మక చర్చ:

మొత్తానికి, తెలుగు భాషా సాంకేతికత అభివృద్ధి వేగంగా జరుగుతున్నప్పటికీ, డేటా కొరత, లిపి నిర్మాణం క్లిష్టత, ఉచ్చారణ వైవిధ్యం వంటి సమస్యలు ఇంకా పరిష్కారాల కోసం ఎదురు చూస్తున్నాయి. అయినప్పటికీ ప్రస్తుతం జరుగుతున్న పరిశోధనలు, ప్రభుత్వ–ప్రైవేట్ సహకారం, AI పురోగతి—all these—తెలుగు భాష యొక్క డిజిటల్ భవిష్యత్తును మరింత బలోపేతం చేస్తున్నాయి.

ముగింపు మరియు సూచనలు:

తెలుగు భాషా సాంకేతికత పరిశోధన ప్రస్తుతం అభివృద్ధి చెందుతున్న అత్యంత కీలక శాస్త్రీయ రంగాలలో ఒకటి. ప్రపంచం మొత్తం డిజిటల్ మార్పు దిశగా అడుగులు వేస్తున్న వేళ, తెలుగు వంటి మహత్తర భాషను సాంకేతిక వనరులతో సమన్వయం చేయడం అత్యవసరంగా మారింది. ఈ అధ్యయనం స్పష్టంగా చూపించినట్లుగా, తెలుగు భాషలో ధ్వన్యాత్మక, వ్యాకరణ, పదనిర్మాణ ప్రత్యేకతలు సాంకేతిక అభివృద్ధికి కొంతమేర సవాళ్లను సృష్టించినప్పటికీ, అదే సమయంలో భాష యొక్క నిర్మాణ బలమే పరిశోధనలో విశేష అవకాశాలను తెరిచింది.

యంత్రానువాదం, మాట గుర్తింపు, OCR, NLP మోడళ్ల అభివృద్ధి వంటి రంగాలలో తెలుగు భాషా సాంకేతికత గణనీయమైన పురోగతి సాధించింది. ముఖ్యంగా డీప్ లెర్నింగ్ ఆధారిత విధానాలు తెలుగు టెక్స్ట్‌ను అర్థం చేసుకోవడం, విశ్లేషించడం, ఉత్పత్తి చేయడంలో తక్కువ సమయంలోనే మంచి మెరుగుదలను కనబరిచాయి. అయినప్పటికీ పూర్తి స్థాయి ఖచ్చితత్వం, సామెతల అర్థం, వాడుక భాష, ప్రాదేశిక భాషా రూపాల అర్థవివేకం వంటి అంశాలు ఇంకా సవాల్‌గా నిలుస్తున్నాయి.

సాంకేతికాభివృద్ధి పరంగా చేసిన పరిశీలనలు సూచిస్తున్నట్లుగా, డేటా కొరత, నాణ్యమైన అనోటేషన్ అవసరం,

వనరుల విభిన్నత లేకపోవడం వంటి సవాళ్లు ఇప్పటికీ తెలుగు NLP అభివృద్ధికి ప్రధాన ఆటంకాలుగా ఉన్నాయి. అయినప్పటికీ ప్రభుత్వ–విశ్వవిద్యాలయ–ప్రైవేట్ రంగాల మధ్య జరుగుతున్న సహకారం, AI ఆధారిత పెద్ద భాషా నమూనాల పెరుగుదల, ఓపెన్ సోర్స్ డేటాసెట్‌ల విస్తరణ—all these—భవిష్యత్‌లో తెలుగు భాషా సాంకేతికతకు మరింత బలమైన మార్గాలను సృష్టించగలవు.

మొత్తానికి, తెలుగు భాషను డిజిటల్ ప్రపంచానికి పూర్తిస్థాయిలో అనుసంధానించేందుకు కొనసాగుతున్న ప్రయత్నాలు భవిష్యత్ తరం కోసం భాషా సంరక్షణ, విజ్ఞానం, విద్య, పరిపాలన రంగాలలో అపారమైన అవగాహనను, అవకాశాలను అందించగలవు. తెలుగు భాషా సాంకేతికత కేవలం భాషను డిజిటల్ చేయడం మాత్రమే కాదు—తెలుగు మాట్లాడే కోట్లాది మందికి సమాన అవకాశాలను అందించే మార్గం.

సూచనలు (Recommendations) – విస్తృత రూపం:

తెలుగు భాషా సాంకేతికత అభివృద్ధి చెందడానికి భవిష్యత్‌లో చేపట్టాల్సిన చర్యలు, మార్గదర్శకాలు కింద ఇవ్వబడ్డాయి:

1. భాషా డేటాబేస్‌లు మరియు కోర్పస్ అభివృద్ధి:

- పెద్ద పరిమాణంలో ప్రామాణికమైన, అనోటేట్ చేసిన డేటాసెట్‌లు రూపొందించాలి.
- వివిధ రంగాల నుంచి (వైద్యం, వ్యవసాయం, చట్టం, విద్య, మీడియా) డేటాను సేకరించి డొమైన్-స్పెసిఫిక్ కోర్పస్ తయారు చేయాలి.
- భాషా డేటాను ఓపెన్ సోర్స్‌గా అందుబాటులో ఉంచడం పరిశోధన వేగాన్ని పెంచుతుంది.

2. భాషా గుర్తింపు మరియు యంత్రానువాదం అభివృద్ధికి ప్రత్యేక పరిశోధన:

- తెలుగు కోసం ప్రత్యేకంగా Morphology-based NLP మోడళ్లు రూపొందించాలి.
- Transformer మరియు LLMల ఆధారంగా చిన్న పరిమాణంలో కూడా సమర్థవంతంగా పనిచేసే మోడళ్లపై పరిశోధన చేయాలి.
- సామెతలు, ప్రయోగాలు, ప్రాంతీయ భాషా రూపాల కోసం ప్రత్యేక ఉపకోర్పస్ అవసరం.

3. వాక్య గుర్తింపు (Speech Recognition) మెరుగుదల:

- కోస్తా, తెలంగాణ, రాయలసీమ వంటి ప్రాంతీయ ఉచ్చారణ డేటాలను సేకరించాలి.
- బాలల, మహిళల, వృద్ధుల స్వర నమూనాలను చేర్చడం ద్వారా మోడళ్ల ఖచ్చితత్వం పెరుగుతుంది.
- తక్కువ శబ్దంతో ఉన్న వాతావరణం–అధిక శబ్దంతో

ఉన్న వాతావరణం వంటి పరస్పర భిన్న పరిస్థితులలో డేటా అవసరం.

4. OCR మరియు చేతిరాత గుర్తింపులో పురోగతి:

- గుణింతాలు, సమాస అక్షరాలు, కళా ఫాంట్లు గుర్తించగల Vision Transformer–based OCR అభివృద్ధి చేయాలి.
- పాత గ్రంథాలు, పాలెగాళ్ల రికార్డులు వంటి పాత పత్రాల–గ్రంథాల డిజిటైజేషన్ కోసం ప్రత్యేక OCR వేరియంట్లు అవసరం.
- చేతిరాత అక్షరాల కోసం ప్రత్యేక హ్యాండ్‌రైటింగ్ డేటాబేస్‌ను అభివృద్ధి చేయాలి.

5. శిక్షణ మరియు అకడమిక్ రంగం అభివృద్ధి:

- విశ్వవిద్యాలయాలలో భాషా సాంకేతికత, NLP, Computational Linguistics వంటి కోర్సులు అందుబాటులోకి తేవాలి.
- పరిశోధనకారులకు ప్రత్యేక ఫెలోషిప్‌లు, ప్రాజెక్ట్ గ్రాంట్‌లు ఇవ్వాలి.
- తెలుగు ఉపాధ్యాయులకు సాంకేతిక శిక్షణ అందించడం భాషా విజ్ఞాన వ్యాప్తికి సహకరిస్తుంది.

6. ప్రభుత్వ–ప్రైవేట్ రంగాల సహకారం:

- ఈ-గవర్నెన్స్ సేవలన్నిటిలో తెలుగు భాషా ఇన్‌పుట్, OCR, Speech-to-Text ఉపయోగాన్ని పెంచాలి.
- టెక్ కంపెనీలతో భాగస్వామ్యం ద్వారా తెలుగు Conversational AI తయారీ వేగవంతమవుతుంది.
- భాషా సాంకేతిక పరిశోధన కోసం ప్రత్యేక తెలుగు AI మిషన్ ఏర్పాటు చేయాలి.

7. సామాజిక, విద్యా వినియోగం పెంపు:

- పాఠశాల–కళాశాల విద్యార్థులకు తెలుగు AI ఆధారిత లెర్నింగ్ టూల్స్ అభివృద్ధి చేయాలి.
- దృష్టి, వినికిడి సమస్యలుగల వ్యక్తులకు Screen Readers, Voice Assistants వంటి సాధనాలను తెలుగు కోసం విస్తరించాలి.
- గ్రామీణ ప్రాంతాల్లో డిజిటల్ అక్షరాస్యత కోసం తెలుగు భాషా సాంకేతిక పరిష్కారాలను అమలు చేయాలి.

సూచనలు (References in Telugu):

1. పుస్తకాలు:

1. భద్రిరాజు కృష్ణమూర్తి. *ద్రావిడ భాషలు.* కేంబ్రిడ్జ్ యూనివర్శిటీ ప్రెస్, 2003.
2. వేదం వెంకటరాయ శాస్త్రి. *తెలుగు భాషా చరిత్ర.* ఎమ్మెస్కో పబ్లికేషన్స్, 2010.
3. శ్రీహరి కె. *కంప్యూటేషనల్ లింగ్విస్టిక్స్ మరియు భారతీయ భాషలు.* యూనివర్శిటీస్ ప్రెస్, 2015.

4. ప్రసాద్ బి. *ఇండియన్ లాంగ్వేజ్ ప్రాసెసింగ్ – సిద్ధాంతం మరియు ప్రయోగాలు.* పిహెచ్ఐ లెర్నింగ్, 2014.
5. సుబ్బారావు కె.వి. *దక్షిణాసియా భాషల వ్యాకరణ తులనాత్మక అధ్యయనం.* కేంబ్రిడ్జ్ యూనివర్శిటీ ప్రెస్, 2012.

2. పరిశోధన వ్యాసాలు (Journal Articles):

6. రెడ్డి ఎస్., శారాఫ్ ఎస్. “భారతీయ భాషల కోసం క్రాస్-లాంగ్వేజ్ వర్డ్ ఎంబెడ్డింగ్స్.” *లాంగ్వేజ్ టెక్నాలజీ అండ్ కంప్యూటేషన్ జర్నల్*, 2018.
7. మూర్తి కె.ఎన్., రావు పి. “డీప్ న్యూరల్ నెట్‌వర్క్ ఆధారంగా తెలుగు వాక్య గుర్తింపు.” *ఇంటర్నేషనల్ జర్నల్ ఆఫ్ స్పీచ్ టెక్నాలజీ*, 2019.
8. లలితాదేవి ఎస్. “భారతీయ భాషల కోసం NLP సాధన అభివృద్ధిలోని సవాళ్లు.” *కంప్యూటేషనల్ లింగ్విస్టిక్స్ జర్నల్*, 2017.
9. రమేష్ కె. “తెలుగు–ఇంగ్లీష్ యంత్రానువాద విధానాల పరిశీలన.” *ఇండియన్ లింగ్విస్టిక్స్ రివ్యూ*, 2020.
10. అనూష గి. “తెలుగు లిపి OCR సాంకేతికత: సమస్యలు మరియు పురోగతి.” *ఇంటర్నేషనల్ జర్నల్ ఆఫ్ ప్యాటర్న్ రికగ్నిషన్*, 2021.

3. కాన్ఫరెన్స్ పేపర్లు:

11. అభిలాష్ బి. మరియు ఇతరులు. “తెలుగు కోసం న్యూరల్ మెషిన్ ట్రాన్స్‌లేషన్ నమూనా.” *ఐకాన్—ఇంటర్నేషనల్ కాన్ఫరెన్స్ ఆన్ NLP*, 2019.
12. ప్రసాద్ ఆర్., శ్రీలత ఎన్. “సీఎన్ఎన్ ఆధారంగా తెలుగు చేతిరాత అక్షరాల గుర్తింపు.” *ఐఈఈఈ డాక్యుమెంట్ అనాలిసిస్ కాన్ఫరెన్స్*, 2020.
13. రాజశేఖర్ ఎన్. “CRF నమూనాలతో తెలుగు పదభాగ (POS) గుర్తింపు.” *ACL – ఇండియన్ లాంగ్వేజ్ NLP వర్క్‌షాప్*, 2018.
14. హరీష్ కుమార్. “ద్రావిడ భాషల కోసం మాట–పాఠ్య మార్పిడి వ్యవస్థలు.” *ICASS పేపర్ సిరీస్*, 2021.

4. ప్రభుత్వ సంస్థల వనరులు:

15. టిడిల్ (Indian Language Technology Program). “తెలుగు భాషా వనరులు మరియు NLP సాధనాలు.” భారత ప్రభుత్వం, ఎలక్ట్రానిక్స్–ఐటి మంత్రిత్వశాఖ, 2020.
16. IIIT హైదరాబాద్ – LTRC. “తెలుగు భాషా సాంకేతికతపై పరిశోధన నివేదికలు.” IIIT Hyderabad Research Repository, 2015–2023.
17. సీడాక్ హైదరాబాద్. “భారతీయ భాషల ధ్వని–పాఠ్య (Speech &

OCR) డేటాసెట్‌లు." CDAC Publications, 2019.

18. పొట్టి శ్రీరాములు తెలుగు విశ్వవిద్యాలయం. "తెలుగు నిఘంటువులు మరియు కోర్పస్ అభివృద్ధి ప్రాజెక్టులు." తెలుగు విశ్వవిద్యాలయ నివేదికలు, 2018.

5. ఆన్లైన్ వనరులు:

19. గూగుల్ రీసెర్చ్. "భారతీయ భాషల బహుభాషా మాట గుర్తింపు." గూగుల్ AI బ్లాగ్, 2020.
20. మైక్రోసాఫ్ట్ రీసెర్చ్ ఇండియా. "భారతీయ భాషల కోసం AI & NLP పరిశోధన." MSR Publications, 2021.
21. ఇండిక్ NLP కన్సార్టియం. "తెలుగు NLP లైబ్రరీ & కోర్పస్ వనరులు." 2019.
22. వికీపీడియా రచయితలు. "తెలుగు భాషా సాంకేతికత." *వికీపీడియా – ఉచిత విజ్ఞానసర్వస్వం*, 2024.

6. థీసిస్‌లు / పరిశోధనా గ్రంథాలు:

23. రావు ఎం. "తెలుగు రూప నిర్మాణ (Morphology) కోసం NLP సాధనాల అభివృద్ధి." పీహెచ్‌డీ థీసిస్, హైదరాబాద్ విశ్వవిద్యాలయం, 2016.
24. అనిల్ కుమార్ కె. "తెలుగు లిపి OCR వ్యవస్థలపై అధ్యయనం." ఎం.టెక్ థీసిస్, జేఎన్‌టియు హైదరాబాద్, 2020.
25. కావిత గి. "హైబ్రిడ్ మోడళ్లతో తెలుగు మాట–సంశ్లేషణ (Speech Synthesis)." పీహెచ్‌డీ పరిశోధన గ్రంథం, IIIT హైదరాబాద్, 2019.