



বাংলা ভাষা

প্রযুক্তির সম্ভাবনা

মামুন অর রশীদ

ভাষা প্রযুক্তি বিশেষজ্ঞ, বাংলাদেশ কমপিউটার কাউন্সিল ও সহযোগী অধ্যাপক, জাহাঙ্গীরনগর বিশ্ববিদ্যালয়

বাংলাকে জাতিসঙ্ঘের ভাষা করতে হলে কী করতে হবে? অথবা সোফিয়া বা আসিমোর মতো রোবটগুলো কি বাংলায় কথা বলতে পারবে? সিরি বা করটানায় বাংলা যুক্ত হতে দেরি হচ্ছে কেন? এসব প্রশ্নের উত্তর কিন্তু এক সূত্রে গাথা। তা হলো, কৃত্রিম বুদ্ধিমত্তাকেন্দ্রিক প্রযুক্তিতে বাংলা ভাষার সন্নিবেশ করার জন্য যথেষ্ট রিসোর্স এখনও নেই।

এ কথা ঠিক, ব্যবহারের দিক থেকে বাংলা প্রথম দিকে থাকা ভাষাগুলোর একটি। বাংলা ভাষাভাষীর রয়েছে রক্তস্নাত ভাষা-আন্দোলনের ইতিহাস। দেশ ও ভাষার মর্যাদা রক্ষায় এই জাতির রয়েছে গৌরবময় ঐতিহ্য, রয়েছে ভাষার প্রতি দরদ, ভাষাকে সম্মুন্নত রাখার চেতনা। কিন্তু দুঃখজনক হলেও সত্য, বাংলা ভাষাকে প্রযুক্তিবান্ধব করার ক্ষেত্রে প্রয়োজনীয় ভিত্তি তৈরি হয়নি, বিশেষ করে কমপিউটিংয়ে বাংলা ভাষাকে অভিযোজিত করার ক্ষেত্রে— খুব বেশি অগ্রসর হয়নি। দুঃখজনক হলেও সত্য, বাংলা এখনও লো-রিসোর্স ল্যাঙ্গুয়েজ।

বাংলা ভাষানির্ভর সফটওয়্যার তৈরির জন্য প্রয়োজনীয় ডাটা বা কর্পাসের পরিমাণ এখনও যথেষ্ট নেই। অথচ বাজারের দিক থেকে চিন্তা করলে কিংবা দেশি ভাষা বিকারের জিও-পলিটিক্যাল গুরুত্বের কথা চিন্তা করলেও বাংলা নিয়ে প্রচুর কাজ করা উচিত। এই কাজ বলতে যেমন, তাত্ত্বিক গবেষণা বোঝাচ্ছি তেমনি, বিনিয়োগ করে বাংলা ভাষার সফটওয়্যার তৈরির কথাও বলছি।

কেন বাংলা নিয়ে কাজ করা উচিত

বাংলা নিয়ে কাজ করার কারণ অনেকগুলো। প্রথমত, স্থানীয়করণ বা লোকালাইজেশন করার জন্য। আগামী দিনের পৃথিবী কেবলই ইংরেজিনির্ভর থাকবে না। লোকাল ভাষাগুলোও প্রযুক্তিজগতে স্থান করে নেবে। যত বেশি স্থানীয় ভাষায় প্রযুক্তি তৈরি হয়, জনসম্পৃক্ততা তত বেড়ে যায়। ভাষা যেন প্রযুক্তি ব্যবহারের প্রতিবন্ধকতা তৈরি না করতে পারে এ জন্যই লোকালাইজেশন। ভাষা কোনো নিরীহ ব্যাপার নয়। রফতানিযোগ্য কোনো দ্রব্যের গায়ে কোন ভাষা ব্যবহার হবে, তা আসলে ব্যবহার করে সেই ভাষার দেশের অর্থনৈতিক ভিত্তির ওপর। মানে ভাষার প্রকাশ মানে অর্থনৈতিক শক্তিরও প্রকাশ। এই ভাষার প্রয়োগ শুধু ভোক্তার সংখ্যার ওপর নির্ভর করে না, নির্ভর করে ভোক্তার মানসিকতা, ভাষাটির স্ট্যাটাস ও প্রযুক্তিগত অবস্থান। কোনো ভাষার লিপি ইউনিকোড সাপোর্ট না করলে তাকে সাইবার জগতে প্রতিষ্ঠিত করা কঠিন। অর্থাৎ প্রযুক্তিগত সমর্থন থাকায় ভাষাটির ব্যবহার বেড়েছে। যেমন নোকিয়া ফোনে বাংলা ভাষাবান্ধব করার ফলে এর গ্রাহক বেড়েছিল।

বাংলা নিয়ে কাজ করার প্রধান কারণ হলো অর্থনৈতিক গুরুত্ব। একে বাজারি গুরুত্বও বলা যেতে পারে। আগামী দশকে দুনিয়া জুড়ে মিলিয়ন মিলিয়ন ডলারের ভাষা-প্রযুক্তিবিষয়ক কাজ করা হবে। গবেষণার ফল এমনটিই জানিয়েছে। ট্রাকটিকার গবেষণা অনুসারে, আগামী পঁচিশ সালের মধ্যে এর বাজার ২০ হাজার মিলিয়ন ডলার ছাড়িয়ে যাবে। ভাষা-প্রযুক্তির একটি অংশ এনএলপি। এই পরিমাণ শুধু এনএলপির জন্য প্রয়োজ্য। এর মধ্যে রয়েছে টেক্সট প্রসেসিং, ইমেজ প্রসেসিং, স্পিচ প্রসেসিং। টেক্সট প্রসেসিংয়ের বড় উদাহরণ গুগল সার্চ ইঞ্জিন। আমরা যদি শুধু ধান... কম্পোজ করলে যে ইঞ্জিন 'ধানমন্ডি' সাজেশন দেয়, এটা হলো ইন্টেলিজেন্ট টেক্সট প্রসেসিংয়ের ফল। নেস্টেড ওয়ার্ড সাজেশন দেয়া হয় কোলোকেশন ও এন-গ্রাম ফিচার ব্যবহার করে। টেক্সট প্রসেসিংয়ের আরেকটি বড় প্রয়োগ মেশিন ট্রান্সলেশন। ইমেজ প্রসেসিংয়ের প্রয়োগ বলা যায় ফেসবুকের অটোটি্যাগিং ফিচারটিকে। এর অর্থ স্বয়ংক্রিয়ভাবে ফেসবুক কোনো ব্যক্তির ছবি চিনতে পারে। বেশিরভাগ ক্ষেত্রেই তা নির্ভুল হয়। ইমেজ প্রসেসিংয়ের আরেকটি প্রয়োগ গাড়ির নম্বর প্লেট ডিটেকশন। স্পিচ প্রসেসিংয়ের ভালো উদাহরণ হলো সিরি, করটানা বা অ্যালেক্সার মতো বুদ্ধিমান মেশিনগুলোর স্পিচ রিকগনিশন অংশটি। বাংলা ভাষার জন্য এই প্রসেসিংগুলো করা হলে দেশের অনেকগুলো সার্ভিস বদলে যাবে। সত্যিকারের পরিবর্তনের ছোঁয়া আসবে।

ভাষা-প্রযুক্তির প্রধান জানালা কৃত্রিম বুদ্ধিমত্তা। ভাষা-প্রযুক্তি ও কৃত্রিম বুদ্ধিমত্তা দুটি ভিন্ন বিষয়। কিন্তু এর মধ্যে সম্পর্কের সেতু গড়ে তৈরি হয় কথা বলা রোবটের মতো রিয়েল লাইফ অ্যাপ্লিকেশন তৈরি করার সময়। ভবিষ্যতের পৃথিবী রোবটিক্সের বা কৃত্রিম বুদ্ধিমত্তার পৃথিবী। ভাষাকেন্দ্রিক এআই প্রযুক্তির প্রধান একটি খাত হলো এনএলপি। ভাষাকে গাণিতিক প্রক্রিয়ায় বিশ্লেষণ করে অ্যাপ্লিকেশন তৈরির উপাদান প্রস্তুত করার বিদ্যাকে বলা হয় এনএলপি বা ন্যাচারাল ল্যাঙ্গুয়েজ প্রসেসিং। এ ধরনের বিদ্যার সাথে অনেকগুলো প্রযুক্তি জড়িয়ে

কাজ চলমান রয়েছে (প্রতি পর্বে এক বা একাধিক প্রতিষ্ঠানের উদ্যোগ নিয়ে বলার চেষ্টা করব। এই পর্বে বিসিসির প্রকল্পের কথা বলা সম্ভব মনে করছি)। এই প্রকল্পটির উদ্দেশ্য বাংলা ভাষার জন্য বিভিন্ন প্রযুক্তিমাধ্যমে (ওয়েব, মোবাইল, কমপিউটার) ব্যবহারযোগ্য বিভিন্ন সফটওয়্যার/টুল/রিসোর্স উন্নয়ন করা, যাতে বাংলা ভাষা কমপিউটারে ব্যবহার করতে কোনো প্রতিবন্ধকতা না থাকে। এর মাধ্যমে বাংলা ভাষার জন্য ১৬টি সফটওয়্যার/টুল/রিসোর্স উন্নয়ন করা হচ্ছে। প্রধান কম্পোনেন্টগুলো হলো— বাংলা কর্পাস উন্নয়ন, বাংলা থেকে

প্রকাশনা শিল্পে ও ওয়েব মাধ্যমে বাংলা ভাষা সঠিকভাবে ব্যবহারের সহজ পথ তৈরি হবে। এই বানান ও ব্যাকরণ পরীক্ষক সফটওয়্যারটি বাংলা একাডেমির প্রমিত বানানবিধি ও প্রমিত বানান অভিধানকে অনুসরণ করবে।

যান্ত্রিক অনুবাদের মাধ্যমে বাংলা ভাষাকে বিভিন্ন ভাষায় স্বয়ংক্রিয়ভাবে অনুবাদ করা যায়। এর অনুবাদকের মাধ্যমে তথ্যমূলক বাংলা, দৈনন্দিন বাংলা, প্রাতিষ্ঠানিক রচনা/ডকুমেন্টস/নথি, সংবাদ বিজ্ঞপ্তি, আবহাওয়া সংবাদ দ্রুত নির্ভুলভাবে অনুবাদ করা সম্ভব হবে। এই অনুবাদকের মাধ্যমে বাংলা থেকে ইংরেজি এবং ইংরেজি থেকে বাংলা ছাড়াও বাংলা থেকে স্প্যানিশ, ফরাসি, জার্মান, রুশ, মান্দারিন, জাপানিজ, কোরিয়ান, আরবি, হিন্দি ভাষায় এবং উল্লিখিত ভাষাগুলোকে থেকে বাংলায় অনুবাদ করা যাবে।

ওসিআরের মাধ্যমে কমপিউটার কম্পোজ করা বাংলা, টাইপরাইটারে মুদ্রিত বাংলা এবং লেটার প্রেসে মুদ্রিত বাংলা ডকুমেন্টকে স্ক্যান করে বা ছবি তোলা মাধ্যমে সার্ভেল ক্যারেক্টারে রূপান্তরিত করা যাবে। এই কম্পোনেন্ট বাংলা হাতের লেখা শনাক্ত করতে পারবে। এর মাধ্যমে বাংলা হাতের লেখা রয়েছে এমন একটি অপরিবর্তনযোগ্য ফাইল বা ছবিকে সিলেক্ট করার পর তাতে অঙ্কিত হাতের লেখা ইউনিকোডে রূপান্তরিত হবে। স্টাইলাস বা আঙুল দিয়ে প্যালেটে বা মোবাইল স্ক্রিনে বর্ণ বা শব্দ লেখার পর তা বাংলা ইউনিকোড টেক্সটে রূপান্তরিত হবে।

রেকর্ড করা বা চলমান বাংলা কথাকে লেখায় রূপান্তর করে স্পিচ টু টেক্সট (STT) সফটওয়্যার। এই অ্যাপ্লিকেশনটি সম্পন্ন হলে বাংলা ভাষার ভাষণ ও বক্তব্য দ্রুত লিখিত বা কম্পোজ অবস্থায় পাওয়া যাবে। বিভিন্ন সাক্ষাৎকার, বিবৃতি দ্রুত যন্ত্রের মাধ্যমে অনুলিখন করা যাবে, যার ফলে অনেক অর্থ-সময় ও শ্রম বাঁচবে। পক্ষান্তরে টেক্সট টু স্পিচ অ্যাপ্লিকেশন হলো ডিজিটাল টেক্সটকে উচ্চারিত শব্দে রূপান্তর করা। এই অ্যাপ্লিকেশন যাদের ডিসলেক্সিয়া (dyslexia), অর্থাৎ পড়ার অসুবিধা বা দৃষ্টি-বৈকল্য আছে, তাদের উপকারে আসবে। এর ফলে স্বয়ংক্রিয়ভাবে যন্ত্রের মাধ্যমে সরকারি জরুরি বিজ্ঞপ্তি, নির্দেশনা, পত্রিকার শিরোনাম/তাজা খবর শোনা যাবে। ওয়েবসাইটে প্রকাশিত লেখা সহজে শোনা যাবে।

বাংলা ভাষাকে রিসোর্সফুল ভাষায় উন্নীত করার প্রধান উপায় হলো করপোরা তৈরি। বাংলায় এজন্য তৈরি হচ্ছে একটি বিশাল করপাস। করপাসটি হবে প্রতিনিধিত্বমূলক এবং ব্যালেন্সড। ভাষা-প্রযুক্তির ভাষায় এটি মূলত অ্যানোটোটেড সিনট্যাকটিক ট্রি-ব্যাংক করপাস হবে। যার অন্তত শতকরা ১০ ভাগ গোল্ড স্ট্যান্ডার্ড রক্ষা করবে, বাকি ৯০ ভাগ সিলভার স্ট্যান্ডার্ড রক্ষা করবে। করপাসটিতে প্রতিনিয়ত নতুন ডাটা যুক্ত হওয়ার সুযোগ থাকবে। ফলে এটি দ্রুত বিশাল ভাণ্ডারে পরিণত হবে।

স্ক্রিন রিডার সফটওয়্যারের মাধ্যমে দৃষ্টিপ্রতিবন্ধী ব্যক্তি বা স্বল্প দৃষ্টিসম্পন্ন ব্যক্তির কমপিউটার বা মোবাইল ব্যবহার করতে পারবেন। কমপিউটারের পর্দায় ভেসে আসা প্যারাফ্রাফের বাংলা লেখা পড়ে শোনাতে এই সফটওয়্যার। কমপিউটারের ▶



আছে। ইদানিং প্রচুর এই শব্দগুলো উচ্চারিত হয়। যেমন— মেশিন লার্নিং, বিগ ডাটা অ্যানালাইসিস, আইওটি প্রভৃতি। এই ভাষা-প্রযুক্তির সাথে কৃত্রিম বুদ্ধিমত্তার সমন্বয়ের কাজটি আমরা দেরিতে অনুধাবন করেছি। বিশেষ করে বাংলা নিয়ে কাজ আমাদের একাডেমিগুলোও করেনি, ইন্ডাস্ট্রিগুলোও করেনি। আজকে ব্যবহারযোগ্য যে প্রযুক্তি পাচ্ছি তার প্রায় সবই পশ্চিমা বিশ্বের প্রস্তুত।

বাংলাদেশে বাংলা ভাষা-প্রযুক্তি নিয়ে কী কাজ হচ্ছে?

সরকারি-বেসরকারি প্রতিষ্ঠান ও বিশ্ববিদ্যালয় পর্যায়ে বাংলা ভাষা নিয়ে সফটওয়্যার উন্নয়ন করা হচ্ছে। বিশেষ করে শাহজালাল বিজ্ঞান ও প্রযুক্তি বিশ্ববিদ্যালয়, বাংলাদেশ প্রকৌশল বিশ্ববিদ্যালয়, ঢাকা বিশ্ববিদ্যালয়, ব্র্যাক বিশ্ববিদ্যালয়, খুলনা প্রকৌশল বিশ্ববিদ্যালয় ভাষা-প্রযুক্তি নিয়ে কাজ করেছে। বিশ্ববিদ্যালয়ের বাইরে অ্যাকসেস টু ইনফরমেশন, বাংলাদেশ কমপিউটার কাউন্সিল, নির্বাচন কমিশন, আন্তর্জাতিক মাতৃভাষা ইনস্টিটিউট, এশিয়াটিক সোসাইটি প্রভৃতি প্রতিষ্ঠান ভাষা-প্রযুক্তি নিয়ে কাজ করছে। বাংলা ডোমেইনকে ভাষা-প্রযুক্তির বিবেচনায় অন্তর্ভুক্ত করলে বিটিআরসি ও বিটিসিএল এ নিয়ে কাজ করছে।

বাংলাদেশ কমপিউটার কাউন্সিলের তত্ত্বাবধানে 'গবেষণা ও উন্নয়নের মাধ্যমে তথ্যপ্রযুক্তিতে বাংলা ভাষা সমৃদ্ধকরণ' প্রকল্পের

পৃথিবীর প্রধান দশটি ভাষায় অটোমেটিক যান্ত্রিক অনুবাদক উন্নয়ন, বাংলা ওসিআর উন্নয়ন (টাইপ করা ও হাতের লেখা অটোমেটিক শনাক্তকরণ ও কম্পোজ), কথা থেকে লেখা এবং লেখা থেকে কথায় রূপান্তর সফটওয়্যার, জাতীয় কিবোর্ডের (বাংলা) উন্নয়ন, বাংলা ফন্ট রূপান্তর ইঞ্জিন, বাংলা বানান ও ব্যাকরণ সংশোধক উন্নয়ন, স্ক্রিন রিডার সফটওয়্যার উন্নয়ন, বাংলা অনুভূতি বিশ্লেষণের সফটওয়্যার উন্নয়ন, ক্ষুদ্র-নৃগোষ্ঠীর ভাষার জন্য কিবোর্ড উন্নয়ন প্রভৃতি।

এ প্রকল্পের আওতায় বাংলা বানান পরীক্ষক ও ব্যাকরণ সংশোধক উন্নয়ন করা হচ্ছে। বানান পরীক্ষক ও ব্যাকরণ সংশোধক হলো বাংলা ভাষার বা শব্দ, বাক্য স্বয়ংক্রিয়ভাবে সম্পাদনা করার সফটওয়্যার। এই সফটওয়্যার শুধু ভুল বানান চিহ্নিত করবে তা নয়, বরং স্বয়ংক্রিয়ভাবে সংশোধনের পরামর্শ দেবে। এই সফটওয়্যারের বানান পরীক্ষক অংশটি মোবাইল, কমপিউটার, ওয়েবসহ অন্যান্য মাধ্যমে বানানসহ ভাষাগত অন্যান্য ভুল চিহ্নিত করবে এবং সঠিক বানান ব্যবহারে পরামর্শ দেবে। বিশেষ করে, একই রকম উচ্চারণ কিন্তু বানান ভিন্ন, একই রকম বানান কিন্তু অর্থ ভিন্ন এমন কনটেক্সটনির্ভর বানান ভুল বিষয়ে সংশোধনী দেবে।

ব্যাকরণ সংশোধক ভুল বাংলা বাক্য জানাতে সাহায্য করবে। সরল ও জটিল বাক্যের প্রচলিত সাধারণ ভুলগুলো চিহ্নিত করে ব্যবহারকারীর কাছে বিকল্পসহ সঠিক বাক্য উপস্থাপন করবে। বানান ও ব্যাকরণ পরীক্ষক প্রফ রিডারের কাজ করবে, যা দ্রুত নির্ভুল রচনা নিশ্চিত করবে। এতে

ইন্টারফেসে থাকা বাটন বা আইকন বাংলা ভাষায় চিনিয়ে দেবে, কমান্ড দেয়া যাবে, যার মাধ্যমে দৃষ্টিহীন ব্যক্তি সহজে কমপিউটার বা মোবাইল ব্যবহার করতে পারবে।

বাক ও শ্রবণ প্রতিবন্ধী ব্যক্তিদের জন্য সাইন টু স্পিচ সফটওয়্যার উন্নয়ন করা হচ্ছে। এর মাধ্যমে একজন ব্যবহারকারী কোনো স্মার্টফোন, ট্যাবলেট বা ক্যামেরায়ুক্ত পিসির সামনে দুই হাত, মুখ ও শরীরের উর্ধ্বাংশের সমন্বয়ে ইঙ্গিত ভাষা প্রকাশ করবেন। সফটওয়্যার এই ইঙ্গিত ভাষাকে বাংলা ইউনিকোড টেক্সটে রূপান্তর করবে। প্রয়োজনে এই টেক্সট সাথে সাথে উচ্চারিত কথায় রূপান্তরিত হবে। ফলে বাক ও শ্রবণ প্রতিবন্ধীদের প্রাথমিক যোগাযোগের সীমাবদ্ধতা দূর করা সম্ভব হবে।

সেন্টিমেন্ট অ্যানালাইসিস টুল সাধারণত কোনো ডকুমেন্ট বা প্যারাগ্রাফের টেক্সট বিশ্লেষণ করে বলতে পারে প্যারাগ্রাফটির বক্তব্য ইতিবাচক/পজিটিভ, নেতিবাচক/নেগেটিভ না নিরপেক্ষ/নিউট্রাল। এর মাধ্যমে ওয়েবসাইটের মন্তব্য ও প্রতিক্রিয়া/ফিডব্যাক বিশ্লেষণ করা যায়। এর মাধ্যমে দ্রুত বাজার-জরিপ, জনমত-জরিপ করা, নির্বাচন-উত্তর জনমত যাচাই যন্ত্রের মাধ্যমে দ্রুত করা যাবে।

বাংলাদেশের ক্ষুদ্র-নগোষ্ঠীর ভাষাগুলোর অধিকাংশ খুব স্বল্প পরিসরে তথ্যপ্রযুক্তির জগতে ব্যবহার হয়। অধিকাংশ ক্ষেত্রেই ভাষার মানসম্পন্ন ডকুমেন্টেশন, রিসোর্স ও ম্যাটেরিয়াল নেই। এদের মধ্যে কয়েকটি বিপন্ন অবস্থায় রয়েছে। অনেক ভাষার পর্যাপ্ত ডিজিটাল ডাটা নেই, ফন্ট ও এনকোডিং নেই। অনেক ভাষার লিপিও নেই। ডিজিটাল রিসোর্স তৈরির মাধ্যমে এই ভাষাগুলোকে প্রযুক্তি জগতে ব্যবহারের উপযোগী করা সম্ভব। এ লক্ষ্যে যেসব ভাষার নিজস্ব বর্ণমালা রয়েছে, সেসব ভাষার জন্য জাতীয় মান অনুসরণ করে কিবোর্ড সফটওয়্যার উন্নয়ন করা হচ্ছে। একই সাথে অন্যান্য বিপন্ন ভাষাগুলোকে ডিজিটাল আর্কাইভে সংরক্ষণের ব্যবস্থা করা হচ্ছে।

ইউনিকোড হলো লিখন পদ্ধতি বা লিপি-পসমূহের কমপিউটারে ব্যবহার করার জন্য আন্তর্জাতিকভাবে স্বীকৃত একটি কমন মান। পৃথিবীর প্রায় সব প্রধান ভাষার লিপির জন্য কমপিউটার ইন্ডাস্ট্রির বিজ্ঞানী, বিভিন্ন দেশের প্রতিনিধি ও বিজ্ঞানীদের সমন্বয়ে গঠিত কনসোর্টিয়াম এই মান নির্ধারণ করে থাকে। এই কম্পোনেন্টের মাধ্যমে বাংলাদেশের পক্ষে বাংলা ভাষা ও অন্যান্য ভাষার ইউনিকোড মান নির্ধারণে সব ধরনের সমর্থন দেয়া হবে। বাংলা ছাড়াও বাংলাদেশের অন্যান্য ভাষার ইউনিকোড মানের সম্ভাব্য যাচাই, প্রণয়ন ও প্রকাশ করা হবে এই কম্পোনেন্টের মাধ্যমে। ইউনিকোড কনসোর্টিয়ামের আরেকটি ইউনিট হলো কমন লোকাল ডাটা রিপোজিটরি (সিএলডিআর)। এটি স্থানীয় ইউনিকোড বিষয়ে বৃহত্তম ও প্রমিত তথ্যভাণ্ডার। আন্তর্জাতিক কোম্পানিগুলো তাদের সফটওয়্যার আন্তর্জাতিকায়ন ও স্থানীয়করণে এই তথ্যভাণ্ডার ব্যবহার করে থাকে। বাংলা ভাষার ক্ষেত্রেও এই কম্পোনেন্টের মাধ্যমে সিএলডিআর উন্নয়ন করা হবে। একই সাথে এর মাধ্যমে ইউনিকোড ও সিএলডিআর প্রমিতকরণ, জমাকরণ ও রক্ষণাবেক্ষণের জন্য প্রয়োজনীয় ব্যবস্থা গ্রহণ করছে। বলা যায়, এই সফটওয়্যার ও রিসোর্সগুলোর

বাংলা ইউনিকোডের পরিবর্তন দরকার

	098	099	09A	09B	09C	09D	09E	09F
0	৭	ঐ	ঔ	ৱ	ি	ি	ি	ি
1	৮	ঊ	ঋ	৳	ি	ি	ি	ি
2	৯	ঋ	ঌ	৴	ি	ি	ি	ি
3	০	঍	ড	৵	ি	ি	ি	ি
4	১	ণ	ত	৶	ি	ি	ি	ি
5	২	থ	দ	৷	ি	ি	ি	ি
6	৩	ধ	ন	৸	ি	ি	ি	ি
7	৪	঩	প	৹	ি	ি	ি	ি
8	৫	ফ	ব	৺	ি	ি	ি	ি
9	৬	ভ	ম	৻	ি	ি	ি	ি
A	৭	য	঺	ৼ	ি	ি	ি	ি
B	৮	঻	়	৽	ি	ি	ি	ি
C	৯	ঽ	ি	৾	ি	ি	ি	ি
D	০	ি	ি	৿	ি	ি	ি	ি
E	১	ি	ি	৿	ি	ি	ি	ি
F	২	ি	ি	৿	ি	ি	ি	ি

বাংলা ইউনিকোড চাটে বাংলা ভাষার যথার্থ প্রতিফল ঘটেনি। এর ফলে বাংলা ডোমেইন এর আইডিএন কার্যক্রম জটিলতার মধ্যে পড়েছে। ভাষা প্রযুক্তির এ বিষয়টি খুবই জরুরি এবং তা আমাদের দেশের গণমাধ্যমসহ প্রায় সবার অলক্ষ্যে সম্পন্ন হয়ে যাচ্ছে।

বাংলা ভাষা ইউনিকোডের প্রথম সংস্করণ থেকে অন্তর্ভুক্ত রয়েছে। সর্বশেষ সংস্করণ অনুযায়ী, বাংলার লিপির জন্য ৯৬টি কোড পয়েন্ট রয়েছে, যার কোড ব্লকের রেঞ্জ ০৯৮০-০৯ঋঋ পর্যন্ত। এই কোড পয়েন্টগুলোতে প্রধানত বাংলা এবং বাংলাসহ অসমীয়া, সংস্কৃত, পালি ভাষায় ব্যবহৃত ব্যঞ্জনবর্ণ, স্বরবর্ণ, গাণিতিক সংখ্যাসহ বিভিন্ন চিহ্ন অন্তর্ভুক্ত।

ইউনিকোডের প্রথম সংস্করণ ১৯৮৮ সালে প্রকাশ পেলেও ১৯৮৮-২০০৯ সাল পর্যন্ত ইউনিকোড কনসোর্টিয়ামে বাংলাদেশ ছিল না। ফলে বাংলা ইউনিকোড বিষয়ে সিদ্ধান্ত গ্রহণ প্রদান করতো পাশ্চাত্য দেশ ভারত ও অন্য দেশের বিশেষজ্ঞেরা। ফলে প্রথম সংস্করণ থেকেই ইউনিকোডে বাংলা ভাষার প্রাথমিক বৈশিষ্ট্যসমূহ যথাযথভাবে রক্ষিত হয়নি। বিশেষ করে, ড, ঢ, য, ঞ কে বাদ দিয়ে চাট করা হয়, পরবর্তীকালে নোকতায়োগে লেখার পরামর্শ দিয়ে ড, ঢ, য-কে অতিরিক্ত বর্ণ হিসেবে যোগ করা হয়, যা বাংলা ভাষার মৌলিক চরিত্রের সাথে সঙ্গতিপূর্ণ নয়। এমন আরো কিছু অসঙ্গতি রয়েছে ইউনিকোড চাটে।

বিভিন্ন সময়ে বাংলাদেশের পক্ষ থেকে ইউনিকোড কনসোর্টিয়ামে বাংলাভাষার এই বিষয়গুলো উত্থাপিত হয়েছে। ২০১০ সালে ‘ডিজিটাল বাংলাদেশ গড়ার লক্ষ্যে বাংলাদেশ সরকার এই কনসোর্টিয়ামে ভোটিং মেম্বর হিসেবে যোগ দেয়। বাংলা ইউনিকোডের উল্লিখিত অসঙ্গতিগুলো কনসোর্টিয়ামে তুলে ধরার জন্য সরকার ও বিশেষজ্ঞমতের পাশাপাশি জনমত প্রয়োজন।

বাংলা ইউনিকোডের কয়েকটি অসঙ্গতি :

নোকতা

ছকে নোকতা ‘Bengali Sign Nukta’ নামে একটি ক্যারেকটার রয়েছে যার কোড পয়েন্ট ০৯BC। সন্দেহাতীতভাবে বাংলা ভাষার সাথে নোকতার কোনো সম্পর্ক নেই। বাংলা ইউনিকোড ছক থেকে একে সরিয়ে নেয়া প্রয়োজন।

ড ঢ য অতিরিক্ত বর্ণ

বাংলা বর্ণমালার তিনটি আবশ্যিক মৌলিক বর্ণকে অতিরিক্ত ব্যঞ্জন বর্ণ ‘Additional consonants’ হিসেবে দেখা হয়েছে। তিনটি হলো ড/জজঅ, ঢ/জ্জঅ, য/গগঅ এবং কোড পয়েন্ট যথাক্রমে ০৯উঈ, ০৯উউ, ০৯উঋ। আসলে তিনটি বর্ণই বাংলা ভাষায় বহু ব্যবহৃত জরুরি মৌলিক বর্ণ, বিকল্প বর্ণ নয়। ছকে তিনটি বর্ণকে ড+নোকতা, ঢ+ নোকতা, য+ নোকতা দিয়ে লেখার নির্দেশনা রয়েছে। অথচ বাংলা ভাষায় এই তিনটি বর্ণ একক পূর্ণ বর্ণ। ইউনিকোডে এদের পূর্ণ বর্ণ হিসেবে স্থান দেয়া প্রয়োজন।

বাংলা দাঁড়ি নেই, দেবনাগরি দণ্ড আছে

বাংলার প্রচলিত দাঁড়ি ইউনিকোডে ব্যবহার করা যায় না, ব্যবহার করতে হয় দেবনাগরি দণ্ড ‘।’, যার কোড পয়েন্ট ০৯৬৪। বাংলা দাঁড়ির ঘরটি (০৯উঈ) সংরক্ষিত বা অব্যবহৃত রয়েছে। গান ও কবিতায় ব্যবহৃত ‘দুই দাঁড়ি’ এর বাঁকানো রূপটিও যুক্ত করা নেই।

টাকার চিহ্ন ও শিরোনাম

বাংলাদেশের মুদ্রা টাকা। শিরোনামের রূপি পরিচয়টি পরিবর্তন করা প্রয়োজন।

কাজ সম্পন্ন হলে দেশ ও জাতি এর সুফল পাবে। প্রযুক্তিতে বাংলা ভাষা আর কোনো প্রতিবন্ধক হবে না, বরং সহায়ক হবে।

আন্তর্জাতিক পর্যায়ে বাংলা ভাষার প্রয়োগ ও বিস্তৃতির বাস্তবিক ভিত্তি তৈরি হবে, যা হবে এক সত্যিকারের বিপ্লব।