ADA

Accurate Data Annotator

ACCURATE DATA ANNOTATOR

BILINGUAL MIRROR GUIDELINE MANUAL

دليل معايير ترميز البيانات - النسخة المقابلة المتناظرة

Resource Library Asset: ADA-MIRROR-2026

Authoritative Quality Control Blueprint

VERSION B: ENGLISH-ARABIC FACE-TO-FACE EDITION

1. INTRODUCTION TO ADA'S ANNOTATION STANDARDS

Welcome to the **Accurate Data Annotator (ADA)** master production pipeline. In the contemporary AI landscape, highly clean, precise, and accurately labeled data forms the structural baseline for production-ready Large Language Models (LLMs). Our core philosophy dictates that linguistic context always beats isolated keywords.

This manual clarifies the rigorous quality compliance standards required across all text annotation processes executed by ADA teams. Objective consistency and reading comprehension are paramount metrics.

2. GENERAL ANNOTATOR RULES

- **Context Mastery:** Never evaluate keywords in absolute isolation. Analyze paragraph intent and context strings.
- **Objective Framing:** Remove cultural, political, or geographical personal bias completely. Rely strictly on data parameters.
- **Linguistic Vigilance:** Treat spelling variations and grammatical structural mutations with strict systematic formatting.

1. مقدمة إلى معايير الترميز لمنصة ADA

مرحباً بكم في خط الإنتاج الرئيسي لمنصة **مبرمجي البيانات (ADA)**. في مشهد الذكاء الاصطناعي المعاصر، تشكّل البيانات النظيفة والدقيقة والمصنفة بعناية الأساس الهيكلي لبناء نماذج اللغات الكبيرة (LLMs) الجاهزة للاستخدام البرمجي الفعلي. إن فلسفتنا الأساسية تنص دائماً على أن السياق اللغوي يتفوق تماماً على الكلمات المفتاحية المنفردة.

يوضح هذا الدليل معايير الامتثال الصارمة للجودة المطلوبة عبر جميع عمليات ترميز وتصنيف النصوص التي تنفذها فرق ADA. يعد الاتساق الموضوعي والاستيعاب القرائي من أهم مقاييس الأداء لدينا.

2. القواعد العامة لمتخصصي الترميز

- **إتقان السياق:** لا تقم أبداً بتقييم الكلمات المفتاحية بشكل معزول ومستقل. قم بتحليل قصد النص بالكامل وسياق العبارات الحقيقي.
- **التأطير الموضوعي:** تخلص تماماً من الانحيازات الشخصية، الثقافية، السياسية أو الجغرافية. اعتمد حصرياً على معايير البيانات.
- **اليقظة اللغوية:** تعامل مع الاختلافات الإملائية والتحورات النحوية الهيكلية بأسلوب تنسيق نظامي صارم وثابت.

3. TEXT ANNOTATION GUIDELINES

3.1 SENTIMENT ANALYSIS

Sentiment analysis categorizes the explicit psychological profile and emotional polarity of a target text block into standard categorical classes:

POLARITY	PRODUCTION PROTOCOL EXAMPLE
POSITIVE	"The batch deployment finished execution ahead of schedule. Exceptional performance."
NEUTRAL	"The system monitoring daemon requested an update sequence check at 14:00."
NEGATIVE	"The storage controller failed completely during replication, causing severe data loss."

3.2 INTENT CLASSIFICATION

Intent maps user queries onto actionable operational logic commands:

INTENT MAPPING RULES

Transactional Intent: Direct user actions requiring account states to change.
Example: "Deactivate my premium server tier instance instantly."

3. إرشادات ترميز وتصنيف النصوص

3.1 تحليل وتصنيف المشاعر

يقوم تحليل المشاعر بتصنيف السمات النفسية الواضحة والقطبية العاطفية لكتلة النص المستهدفة إلى فئات قياسية معتمدة:

القطبية	أمثلة عملية من خطوط الإنتاج
إيجابي	"اكتمل تنفيذ نشر الحزمة البرمجية قبل الجدول الزمني المحدد. أداء استثنائي وممتاز."
محايد	"طلب برنامج مراقبة النظام التحقق من تسلسل التحديث التلقائي في تمام الساعة 14:00."
سلبي	"فشل متحكم التخزين تماماً أثناء عملية النسخ المتطابق، مما تسبب في فقدان فادح للبيانات."

3.2 تصنيف القصد والنية (INTENT)

يقوم تصنيف القصد بربط استعلامات المستخدمين بأوامر برمجية منطقية قابلة للتنفيذ الفعلي:

قواعد تحديد القصد والنية

قصد إجرائي وعملياتي: إجراءات مباشرة من المستخدم تتطلب تغيير حالة الحساب.
مثال: *"قم بإلغاء تنشيط خادم الفئة الممتازة الخاص بي فوراً وبشكل كلي."*

4. NAMED ENTITY RECOGNITION (NER)

NER enforces isolating exact token sequences that represent physical or conceptual corporate and geographical anchors. Spans must follow precise token boundaries without including outside empty spaces.

ENTITY CLASSIFICATION	FORMATTED TOKEN SEGMENT
PER (Person Name)	"The ultimate platform decision was authorized by **[Batoul H. Hassaballa]**."
ORG (Organization)	"Data sets are validated using the standards of **[Accurate Data Annotator]**."
LOC (Location Area)	"Operational servers and core development are centered in **[Egypt]**."

4. تمييز الكيانات المسمّاة (NER)

يفرض نظام تمييز الكيانات المسمّاة عزل متسلسلات الرموز (Tokens) الدقيقة التي تمثل ركائز جغرافية أو مؤسسية أو مادية ملموسة. يجب أن تتبع النطاقات المحددة حدود الكلمات بدقة بالغة دون تضمين مسافات فارغة خارجية.

التصنيف الهيكلي للكيان	مقطع الكلمات المرمزة والمصنفة
PER (اسم الشخص)	"تم اعتماد القرار النهائي والسياسة العليا للمنصة بواسطة **[بتول ح. حسب الله]**".
ORG (المؤسسة / المنظمة)	"يتم التحقق من صحة مجموعات البيانات باستخدام معايير **[دقة مبرمجي البيانات]**".
LOC (الموقع الجغرافي)	"تتركز خوادم العمليات التشغيلية وفريق التطوير الأساسي في **[مصر]**".