

VAREX: A Benchmark for Multi-Modal Structured Extraction from Documents

Supplementary Material

This supplement provides additional detail and examples supporting the main paper:

- **Appendix A** — Evaluation prompt used for all models.
- **Appendix B** — Complete document examples: a Nested-category form (B.1) and a Table-category form (B.2), each with schema and ground truth.
- **Appendix C** — Schema echo examples: three models receiving the same document, illustrating the compliance failure spectrum.
- **Appendix D** — Resolution robustness: 200 DPI vs. 50 DPI visual comparison with per-field error analysis.
- **Appendix E** — Input modality comparison: Plain Text vs. Spatial Text representations of the same document.

A. Evaluation Prompt. The full evaluation prompt used for all models (except h2oVL Mississippi and NuExtract 2.0, which use their published formats):

```
System: Extract structured data from this document. Return a JSON object matching this
      schema: {schema}
Return null for fields you cannot find.
Return ONLY valid JSON.
Return an instance of the JSON with extracted values, not the schema itself.
```

B. Document Examples. We present two complete examples: a Nested document illustrating `$defs/$ref` schema structure, and a Table document illustrating array-of-objects extraction.

B.1. Nested Example: Housing Terms and Conditions. A Nested-category document illustrating `$defs/$ref` schema structure.



OMB No. 1235-0002
Expires/Se caduca: 09/30/2023

HOUSING TERMS AND CONDITIONS

TÉRMINOS Y CONDICIONES DE LA VIVIENDA

Important Notice to Migrant Agricultural Worker: The Migrant and Seasonal Agricultural Worker Protection Act requires the furnishing of the following information.

Aviso Importante Para Obreros Agrícolas Migratorios: La Ley Para la Protección de Obreros Migratorios y Temporeros exige que se provea la información siguiente.

1. The housing is provided by

1. La vivienda la provee

Name Knox Inc

Nombre _____

Address 41516 Anita Plaza Suite 990

Dirección _____

City & state / Zip code

Ciudad y estado / Código Postal

Jasmineberg, TX 44064

2. Individual(s) in charge

2. Persona(s) encargada(s)

Name Joseph, Garima C.

Nombre _____

Address 41516 Anita Plaza Suite 990

Dirección _____

City & state / Zip code

Ciudad y estado / Código Postal

Jasmineberg, TX 44064

Phone (283) 690-4093

Teléfono _____

3. Mailing address of housing facility

3. Dirección de correo de la vivienda

Address 41516 Anita Plaza Suite 990

Dirección _____

City & state / Zip code

Ciudad y estado / Código Postal

Jasmineberg, TX 44064

Phone (337) 134-7229

Teléfono _____

4. Conditions of occupancy

4. Condiciones de ocupación

Who may live in housing facility No smoking, no pets allowed
Charges made for housing (if none, so state)
Meals provided (if none, so state)
Charges for utilities (if none, so state)
Other charges, if any
Other conditions of occupancy

Quién puede habitar la vivienda
Cargos hechos por proporcionar la vivienda (Si no los hay, declárelo)
Comidas proporcionadas (si no las hay, declárelo)
Cargos por servicios(luz, agua, gas) (si no los hay, declárelo)
Otros cargos, si los hay
Otras condiciones de ocupación

Figure 1. Example VAREX document (Nested category). The schema uses \$defs/\$ref to define reusable nested object types (HousingProvider, IndividualInCharge, HousingFacilityMailingAddress). Only the English-language fields contain fillable widgets; the Spanish translation serves as static context.

Schema:

```
{
  "type": "object",
  "$defs": {
    "HousingFacilityMailingAddress": {
      "type": "object",
      "properties": {
        "address": {"type": "string", "description": "Mailing address of the housing facility"},
        "city_state_zip": {"type": "string", "description": "City, state, and zip code"},
        "phone": {"type": "string", "description": "Phone number"}
      }
    },
    "HousingProvider": {
      "type": "object",
      "properties": {
        "name": {"type": "string", "description": "Name of the housing provider"},
        "address": {"type": "string", "description": "Address of the housing provider"},
        "city_state_zip": {"type": "string", "description": "City, state, and zip code"}
      }
    },
    "IndividualInCharge": {
      "type": "object",
      "properties": {
        "name": {"type": "string", "description": "Name of the individual in charge"},
        "address": {"type": "string", "description": "Address of the individual in charge"},
        "city_state_zip": {"type": "string", "description": "City, state, and zip code"},
        "phone": {"type": "string", "description": "Phone number"}
      }
    }
  },
  "properties": {
    "conditions_of_occupancy": {"type": "string", "description": "Conditions of occupancy"},
    "housing_provider": {"$ref": "#/$defs/HousingProvider"},
    "individual_in_charge": {"$ref": "#/$defs/IndividualInCharge"},
    "housing_facility_mailing_address": {"$ref": "#/$defs/HousingFacilityMailingAddress"}
  }
}
```

Ground Truth:

```
{
  "conditions_of_occupancy": "No smoking, no pets allowed",
  "housing_provider": {
    "name": "Knox Inc",
    "address": "41516 Anita Plaza Suite 990",
    "city_state_zip": "Jasmineberg, TX 44064"
  },
  "individual_in_charge": {
    "name": "Joseph, Garima C.",
    "address": "41516 Anita Plaza Suite 990",
    "city_state_zip": "Jasmineberg, TX 44064",
    "phone": "(283) 690-4093"
  },
  "housing_facility_mailing_address": {
    "address": "41516 Anita Plaza Suite 990",
    "city_state_zip": "Jasmineberg, TX 44064",
    "phone": "(337) 134-7229"
  }
}
```

Each value was programmatically written to a specific PDF widget and is deterministic. The nested structure mirrors the \$defs definitions in the schema.

B.2. Table Example: ALJ Application. A Table-category document illustrating array-of-objects extraction with value truncation.

United States Office of Personnel Management		Form Approved: OMB No. 3206-0248	
Application for Senior Administrative Law Judge			
<p>Description of the Senior Administrative Law Judge (SALJ) Program: The SALJ Program allows retired Administrative Law Judges (ALJs) to be reemployed on a temporary and intermittent basis to complete hearings of one or more specified case(s) in accordance with the Administrative Procedure Act of 1946. Upon appointment, and while reemployed, the retired ALJ is referred to as a SALJ. (See 5 U.S.C. § 3323(b), and 5 CFR 930.201, et seq)</p> <p>Who Can Apply: Federal ALJs who retired from the Federal Government, and who are currently receiving a federal annuity under the Civil Service Retirement System (CSRS) or the Federal Employee Retirement System (FERS). Retired ALJs are required to demonstrate through proper documentation that they are licensed and authorized to practice law under the laws of a State, the District of Columbia, the Commonwealth of Puerto Rico, or other territorial court established under the Constitution. Judicial status is acceptable in lieu of "active" status in States that prohibit sitting judges from maintaining "active" status to practice law. Being in "good standing" is also acceptable in lieu of "active" status in States where the licensing authority considers "good standing" as having a current license to practice law. If eligible, the U.S. Office of Personnel Management (OPM) will place the retired ALJs name on its SALJ Master List for referral to a hiring agency for possible reemployment. Placement on the SALJ Master List is not a guarantee of reemployment.</p> <p>Instructions: Type or print clearly in black or blue ink. Complete the application and Geographic Preference Statement for Senior Administrative Law Judge Applicant (OPM Form 1655-A). Submit your application and requested documentation to: Administrative Law Judge Program Office, U.S. Office of Personnel Management, 1900 E Street, NW, Washington, DC 20415 or Email to ali.application@opm.gov.</p>			
1. Name (Last, First, MI) Renee T. Estrada		2. Last 5 Digits of SSN 09801	3. Telephone Numbers Daytime () 2109652 Other () 2790389
4. Address (Street, City, State, ZIP Code) 1544 Michael Burgs Suite 985			5. Email Address (if available) restrada@campbell.com
6. Are you a U.S. Citizen? <input type="radio"/> Yes <input type="radio"/> No		7. Highest Grade and Step/Level and Rate as ALJ Administrative Law Judge, Grade 15, Step 5, \$145,600	
8. Retirement Date (mm/dd/yyyy) 06/15/2023		9. Are you a Civil Service annuitant? <input type="radio"/> No <input type="radio"/> Yes <input checked="" type="radio"/> Yes	
10. Do you meet the licensure requirements described above in "Who Can Apply?" <input type="radio"/> Yes <input type="radio"/> No		<input type="radio"/> CSRS CSA Claim Number: <input type="radio"/> FERS	
11. Since the date of your original appointment as an ALJ, have you ever been disciplined or cited for breach of ethics or unprofessional conduct by, or been the subject of a complaint to, any court, administrative agency, bar association, disciplinary committee, or other professional group? <input type="radio"/> Yes <input type="radio"/> No		If yes, in which jurisdiction: Please describe the nature and outcome of the proceeding (if needed, please attach a separate sheet to this form): California	
12. List your ALJ experience. If desired, attach a resume to this form.			
From (mm/yyyy)	To (mm/yyyy)	Title	Grade/Level
01/15/2005	06/15/2023	Administrativ	Department
03/10/2000	01/14/2005	Hearings	Social
07/20/1995	03/09/2000	Attorney	Department
13. For which agencies are you interested in conducting hearings? Department of Veterans Affairs, Social Security Administration, Department of Labor			
14. Availability to work: <input type="radio"/> Part Time <input type="radio"/> Full Time		15. Are you willing to travel? <input type="radio"/> Yes <input type="radio"/> No	
5 U.S.C. § 3323(b) 5 CFR 930.201 et seq		Page 1	OPM Form 1655 September 2024

Figure 2. Example VAREX document (Table category). The schema contains both nested objects (Address) and arrays of objects (alj_experience, licensure_states). Ground-truth values are trimmed to match the visible rendered text when PDF widget bounding boxes truncate the content (e.g., Departmen rather than Department), ensuring models are evaluated against what is actually readable.

Schema (condensed, field descriptions omitted):

```
{
  "type": "object",
  "$defs": {
    "ALJExperience": {
      "type": "object",
      "properties": {
        "position": {"type": "string"}, "agency": {"type": "string"},
        "start_date": {"type": "string"}, "end_date": {"type": "string"}
      }
    },
    "Address": {"type": "object", "properties": {"street": {"type": "string"}}},
    "LicensureState": {"type": "object", "properties": {"state": {"type": "string"}}}
  },
  "properties": {
    "applicant_name": {"type": "string"}, "last_five_ssn_digits": {"type": "string"},
    "daytime_phone": {"type": "string"}, "other_phone": {"type": "string"},
    "email": {"type": "string"}, "highest_grade_step_level_rate": {"type": "string"},
    "retirement_date": {"type": "string"}, "interested_agencies": {"type": "string"},
    "address": {"$ref": "#/$defs/Address"},
    "alj_experience": {"type": "array", "items": {"$ref": "#/$defs/ALJExperience"}},
    "licensure_states": {"type": "array", "items": {"$ref": "#/$defs/LicensureState"}}
  }
}
```

Ground Truth:

```
{
  "applicant_name": "Renee T. Estrada",
  "last_five_ssn_digits": "09801",
  "daytime_phone": "2109652",
  "other_phone": "2790389",
  "email": "restrada@campbell.com",
  "highest_grade_step_level_rate": "Administrative Law Judge, Grade 15, Step 5, $145,600",
  "retirement_date": "06/15/2023",
  "interested_agencies": "Department of Veterans Affairs, Social Security Administration, Department of Labor",
  "address": {"street": "1544 Michael Burgs Suite 985"},
  "alj_experience": [
    {"position": "Administrativ", "agency": "Departmen", "start_date": "01/15/2005", "end_date": "06/15/2023"},
    {"position": "Hearings", "agency": "Social", "start_date": "03/10/2000", "end_date": "01/14/2005"},
    {"position": "Attorney", "agency": "Departmen", "start_date": "07/20/1995", "end_date": "03/09/2000"}
  ],
  "licensure_states": [{"state": "California"}]
}
```

C. Schema Echo Examples. Three models receive the document in Fig. 3 with the schema shown below. We show the first lines of each output to illustrate the spectrum of structured output compliance failures described in Sec. 4.4.

State of California—Health and Human Services Agency
California Department of Public Health
Laboratory Field Services

ANNUAL TEST VOLUME OF CALIFORNIA SPECIMENS Out-of-State Laboratory

Indicate the annual volume of tests performed on specimens received from California by specialties or subspecialty.

Name of Laboratory Parker Diagnostic Laboratories	State ID Number LAB-2023-CO-	CLIA 01D2345678
Address (number, street) 7862 Davis Tunnel	City Karenview	State ZIP NJ 80751

SPECIALTY/SUBSPECIALTY	ANNUAL TEST VOLUME	SPECIALTY/SUBSPECIALTY	ANNUAL TEST VOLUME
110 Bacteriology	2,000	010 Histocompatibility	600
115 Mycobacteriology	1,500	400 Hematology	500
120 Mycology	800	510 ABO and Rh Type	1,200
130 Parasitology	500	520 Antibody Detection Transfusion	700
140 Virology	1,200	530 Antibody Detection Non-transfusion	900
210 Syphilis Serology	600	540 Antibody Identification	400
220 General Immunology	700	550 Compatibility Testing	300
310 Routine Chemistry	300	610 Histopathology	200
320 Urinalysis	1,500	620 Oral Pathology	1,000
330 Endocrinology	2,000	630 Cytology	800
340 Toxicology	800	900 Clinical Cytogenetics	500
		Sub-Total:	15,750

List all other tests performed and annual test volume (Use additional sheets if necessary)

Paternity Testing	300	Cardiac Marker Testing	100
Drug Screening	400	Vitamin D Testing	80
Genetic Testing	250	Thyroid Function Testing	120
Allergy Testing	150	Hormone Level Testing	90
		Sub-Total:	1,250
		Total Volume:	17,000

Authorized Signature: _____ Printed Name: Laura J. Romero Date: 0611

LAB 1670S (8/11)

Figure 3. Input document for the schema echo comparison. A laboratory certification form with both flat fields and array-of-objects structures (`test_volumes`, `other_tests`) defined via `$defs/$ref`. Three models receive this document with the schema below; their outputs illustrate the compliance failure spectrum described in Sec. 4.4.

Schema (condensed, field descriptions omitted):

```
{
  "type": "object",
  "$defs": {
    "OtherTest": {"type": "object", "properties": {
      "test_name": {"type": "string"}, "annual_volume": {"type": "string"}},
    "TestVolume": {"type": "object", "properties": {
      "annual_volume": {"type": "string"}}}
  },
  "properties": {
    "laboratory_name": {"type": "string"}, "state_id_number": {"type": "string"},
    "clia": {"type": "string"}, "address": {"type": "string"},
    "city": {"type": "string"}, "state": {"type": "string"},
    "zip_code": {"type": "string"}, "sub_total_other_tests": {"type": "string"},
    "sub_total_specialties": {"type": "string"}, "total_volume": {"type": "string"},
    "printed_name": {"type": "string"}, "signature_date": {"type": "string"},
    "test_volumes": {"type": "array", "items": {"$ref": "#/$defs/TestVolume"}},
    "other_tests": {"type": "array", "items": {"$ref": "#/$defs/OtherTest"}}
  }
}
```

Gemini 2.5 Pro (correct extraction):

```
{
  "laboratory_name": "Parker Diagnostic Laboratories",
  "state_id_number": "LAB-2023-CO-",
  "clia": "01D2345678",
  "address": "7862 Davis Tunnel",
  "city": "Kareview",
  "state": "NJ",
  ...
}
```

Qwen3-VL 2B (envelope echo, values extracted but wrapped in schema metadata):

```
{
  "type": "object",
  "properties": {
    "laboratory_name": "Parker Diagnostic Laboratories",
    "state_id_number": "LAB-2023-CO-",
    "clia": "01D2345678",
    "address": "7862 Davis Tunnel",
    ...
  }
}
```

InternVL3.5 1B (pure schema echo, schema definition reproduced verbatim):

```
{
  "laboratory_name": {
    "type": "string",
    "description": "Name of the laboratory"
  },
  "state_id_number": {
    "type": "string",
    "description": "State ID number of the laboratory"
  },
  ...
}
```

Three model outputs on the same document (Image V input). Gemini 2.5 Pro extracts values correctly. Qwen3-VL 2B extracts the correct values but wraps them in the schema's structural metadata ("type": "object", "properties": { . . . }), producing invalid output. InternVL3.5 1B reproduces the schema definition verbatim with no extracted values. Both

failure modes are triggered by `$defs/$ref` in the input schema.

D. Resolution Robustness Example. Table 4 reports that open models in the 8–17B range lose 38–40 pp when image resolution drops from 200 DPI to 50 DPI, while Gemini models lose under 3.5 pp. We illustrate this with a Table-category document (Figs. 4 and 5): a bond loss application containing 11 array rows of serial numbers, dates, and denominations. At 50 DPI, digits and dates become visually ambiguous, producing the OCR-like errors shown in Tab. 1.

Schema (condensed, field descriptions omitted):

```
{
  "type": "object",
  "$defs": {
    "Bond": {
      "type": "object",
      "properties": {
        "bond_identifier": {"type": "string"},
        "denomination": {"type": "number"},
        "issue_date": {"type": "string"},
        "description": {"type": "string"}
      }
    }
  },
  "properties": {
    "customer_name": {"type": "string"},
    "customer_number": {"type": "string"},
    "film_record_unavailable_reason": {"type": "string"},
    "disappearance_circumstances": {"type": "string"},
    "bonds": {"type": "array", "items": {"$ref": "#/$defs/Bond"}}
  }
}
```

Table 1. Representative extraction errors at 50 DPI on the document in Figs. 4 and 5. All models shown score 100% at 200 DPI. Error types at 50 DPI: digit confusion, character substitution, and missed fields.

Model	Field	Ground Truth	Prediction
Maverick 17B×128E	bonds[1].issue_date	05/22/2012	05/20/2012
	bonds[1].bond_id	BBB 98765432 B	BBB 8765432 B
	customer_number	DPM-2020-16514	DPM 2020-16514
Qwen3-VL 8B	customer_name	Nadia I. Jaimes	Nadia J. Jaimes
	customer_number	DPM-2020-16514	DPM-2020-18514
Ministral 14B	customer_name	Nadia I. Jaimes	Nadia I. James
	bonds[6].issue_date	02/28/2021	02/08/2021
Gemini 2.5 Pro	<i>100% EM at both resolutions</i>		
Gemini 2.5 Flash	<i>100% EM at both resolutions</i>		

For Fiscal Service use only: **Nadia I. Jaimes** Customer Name **DPM-2020-16514** Customer No.

FS Form 2517
 Department of the Treasury
 Bureau of the Fiscal Service
 (Revised August 2022)

**APPLICATION FOR RELIEF DUE TO THE LOSS, THEFT,
 OR DESTRUCTION OF PAID UNITED STATES SAVINGS BONDS OR NOTES**

Read the instructions on page 2 before completing this form.
TYPE OR PRINT ALL INFORMATION

The undersigned hereby affirms that the following-described paid United States Savings Bonds or Savings Notes have disappeared as stated and that the information given is true so far as is known and ascertainable.

Note: "Bonds," as referred to below, includes Savings Notes, when appropriate.

A. Description of bonds:

Serial Number (With denominational prefix and letter suffix)	Issue Date (Month and year)	Amount Paid	Date Paid (Month, day and year)	Registration (Social Security Number, names, including middle names or initials and connecting words, exactly as inscribed on the bonds)
AAA 12345678 A	01/15/2010	1000		Series EE bond issued to Nadia Jaimes
BBB 98765432 B	05/22/2012	500		
CCC 45678912 C	09/18/2015	200		
DDD 78912345 D	03/10/2018	100		
EEE 23456789 E	11/30/2019	50		
FFF 56789123 F	07/14/2020	25		
GGG 89123456 G	02/28/2021	10		
HHH 34567891 H	06/16/2022	5		
III 67891234 I	10/20/2011	1000		
JJJ 91234567 J	04/05/2014	500		
KKK 23456789 K	08/12/2017	200		

B. Were the bonds presented for redemption? Yes No

Were the bonds stamped "Paid"? Yes No

C. State the reason(s) a film record of the bonds is not available:

Natural disaster damaged records

D. State the circumstances surrounding the disappearance of and search for the bonds:

Lost during a house fire, searched thoroughly

Figure 4. VAREX document at standard resolution (200 DPI). A Table-category bond loss application with 11 array rows. All top models achieve 100% EM at this resolution.

For Fiscal Service use only: **Nadia I. Jaimes** Customer Name **DPM-2020-16514** Customer No.

FS Form 2517
 Department of the Treasury
 Bureau of the Fiscal Service
 (Revised August 2022)

**APPLICATION FOR RELIEF DUE TO THE LOSS, THEFT,
 OR DESTRUCTION OF PAID UNITED STATES SAVINGS BONDS OR NOTES**

Read the instructions on page 2 before completing this form.
TYPE OR PRINT ALL INFORMATION

The undersigned hereby affirms that the following-described paid United States Savings Bonds or Savings Notes have disappeared as stated and that the information given is true so far as is known and ascertainable.

Note: "Bonds," as referred to below, includes Savings Notes, when appropriate.

A. Description of bonds:

Serial Number <small>(With denominational prefix and letter suffix)</small>	Issue Date <small>(Month and year)</small>	Amount Paid	Date Paid <small>(Month, day and year)</small>	Registration <small>(Social Security Number, names, including middle names or initials and connecting words, exactly as inscribed on the bonds)</small>
AAA 12345678 A	01/15/2010	1000		Series EE bond issued to Nadia Jaimes
BBB 98765432 B	05/22/2012	500		
CCC 45678912 C	09/18/2015	200		
DDD 78912345 D	03/10/2018	100		
EEE 23456789 E	11/30/2019	50		
FFF 56789123 F	07/14/2020	25		
GGG 89123456 G	02/28/2021	10		
HHH 34567891 H	06/16/2022	5		
III 67891234 I	10/20/2011	1000		
JJJ 91234567 J	04/05/2014	500		
KKK 23456789 K	08/12/2017	200		

- B.** Were the bonds presented for redemption? Yes No
 Were the bonds stamped "Paid"? Yes No

C. State the reason(s) a firm record of the bonds is not available:
 Natural disaster damaged records

D. State the circumstances surrounding the disappearance of and search for the bonds:
 Lost during a house fire, searched thoroughly

Figure 5. Same document at reduced resolution (50 DPI). Serial numbers, dates, and small text become visually ambiguous. Open models in the 8–17B range lose 13–35 pp; Gemini models remain at 100%.

E. Input Modality Comparison: Plain Text vs. Spatial Text. The four VAREX modalities provide controlled variation in what information the model receives. The largest accuracy gain comes from adding spatial layout to plain text (+3–18 pp; Sec. 4.3). Below we show both text representations for the Table-category document in Fig. 2 (Appendix B.2), focusing on the region where the difference is most pronounced.

In practice, Spatial Text can be reconstructed from any OCR engine that provides word-level bounding boxes (including lightweight CPU-based systems) without requiring a vision-language model. This makes layout-preserving text a practical input option even in cost-constrained deployments: a small text-only model receiving Spatial Text can approach the accuracy of a larger VLM receiving the document image (Sec. 4.3, Tab. 3).

```

text_flow.txt
3. Telephone Numbers
Daytime (
)
Other
(
)
4. Address (Street, City, State, ZIP Code)
5. Email Address (if available)
6. Are you a U.S. Citizen?
Yes
No
7. Highest Grade and Step/Level and Rate as ALJ
8. Retirement Date (mm/dd/yyyy)
9. Are you a Civil Service annuitant?
No
Yes
CSRS
FERS
|
CSA Claim Number:
10. Do you meet the licensure requirements described above in "Who Can
Apply?"
Yes
No
If yes, list the state(s) and /or jurisdiction(s)
where you currently meet the licensure
requirements (attach proof of licensure):
11. Since the date of your original appointment as an ALJ, have you ever been
disciplined or cited for breach of ethics or unprofessional conduct by, or
been the subject of a complaint to, any court, administrative agency, bar
association, disciplinary committee, or other professional group?
Yes
No
If yes, in which jurisdiction: Please
describe the nature and outcome of the
proceeding (if needed, please attach a
separate sheet to this form):
12. List your ALJ experience. If desired, attach a resume to this form.
From (mm/yyyy)
To (mm/yyyy)
Title
Grade/Level
Employer
(Agency)
Location
(City/State)
13. For which agencies are you interested in conducting hearings?
14. Availability to work:
Part Time
Full Time
15. Are you willing to travel?
Yes
No
Page 1
Renee T. Estrada
09801
2109652
2790389
1544 Michael Burgs Suite 985
restrada@campbell.com
Administrative Law Judge, Grade 15, Step 5, $145,600
06/15/2023
California
01/15/2005
06/15/2023
Administrativ
Department
03/10/2000
01/14/2005
Hearings
Social
07/20/1995
03/09/2000
Attorney
Department
Department of Veterans Affairs, Social Security Administration, Department of Labor

```

Figure 6. Plain Text (P) representation of the ALJ form (Fig. 2), produced by PyMuPDF `get_text()` in reading order. The form header is cropped for space. Note that table columns are serialized vertically, all “From” dates appear in sequence, then all “To” dates, then all titles. Large models can resolve this ordering, but smaller models struggle: the P → S gain ranges from +3 pp for frontier models to +18 pp at smaller scales (Tab. 3).

text_layout.txt
 United States Office of Personnel Management
 Form Approved:
 OMB No. 3206-0248

Application for Senior Administrative Law Judge

Description of the Senior Administrative Law Judge (SALJ) Program: The SALJ Program allows retired Administrative Law Judges (ALJs) to be reemployed on a temporary and intermittent basis to complete hearings of one or more specified case(s) in accordance with the Administrative Procedure Act of 1946. Upon appointment, and while reemployed, the retired ALJ is referred to as a SALJ. (See 5 U.S.C. § 3323(b), and 5 CFR 930.201, et seq)

Who Can Apply: Federal ALJs who retired from the Federal Government, and who are currently receiving a federal annuity under the Civil Service Retirement System (CSRS) or the Federal Employee Retirement System (FERS). Retired ALJs are required to demonstrate through proper documentation that they are licensed and authorized to practice law under the laws of a State, the District of Columbia, the Commonwealth of Puerto Rico, or other territorial court established under the Constitution. Judicial status is acceptable in lieu of "active" status in States that prohibit sitting judges from maintaining "active" status to practice law. Being in "good standing" is also acceptable in lieu of "active" status in States where the licensing authority considers "good standing" as having a current license to practice law. If eligible, the U.S. Office of Personnel Management (OPM) will place the retired ALJs name on its SALJ Master List for referral to a hiring agency for possible reemployment. Placement on the SALJ Master List is not a guarantee of reemployment.

Instructions: Type or print clearly in black or blue ink. Complete the application and Geographic Preference Statement for Senior Administrative Law Judge Applicant (OPM Form 1655-A). Submit your application and requested documentation to: Administrative Law Judge Program Office, U.S. Office of Personnel Management, 1900 E Street, NW, Washington, DC 20415 or Email to alj.application@opm.gov.

1. Name (Last, First, MI) Renee T. Estrada
 2. Last 5 Digits of SSN 09801
 3. Telephone Numbers
 Daytime () 2109652
 Other () 2790389
 5. Email Address (if available) restrada@campbell.com

4. Address (Street, City, State, ZIP Code)
 1544 Michael Burgs Suite 985

6. Are you a U.S. Citizen? Yes No
 7. Highest Grade and Step/Level and Rate as ALJ
 Administrative Law Judge, Grade 15, Step 5, \$145,600

8. Retirement Date (mm/dd/yyyy) 06/15/2023
 9. Are you a Civil Service annuitant? No Yes
 CSRS FERS
 CSA Claim Number:

10. Do you meet the licensure requirements described above in "Who Can Apply?" Yes No
 If yes, list the state(s) and /or jurisdiction(s) where you currently meet the licensure requirements (attach proof of licensure):
 California

11. Since the date of your original appointment as an ALJ, have you ever been disciplined or cited for breach of ethics or unprofessional conduct by, or been the subject of a complaint to, any court, administrative agency, bar association, disciplinary committee, or other professional group? If yes, in which jurisdiction: Please describe the nature and outcome of the proceeding (if needed, please attach a separate sheet to this form):

Yes No

12. List your ALJ experience. If desired, attach a resume to this form.

From (mm/yyyy)	To (mm/yyyy)	Title	Grade/Level	Employer (Agency)	Location (City/State)
01/15/2005	06/15/2023	Administrativ		Departmen	
03/10/2000	01/14/2005	Hearings		Social	
07/20/1995	03/09/2000	Attorney		Departmen	

13. For which agencies are you interested in conducting hearings?
 Department of Veterans Affairs, Social Security Administration, Department of Labor

14. Availability to work: Part Time Full Time
 5 U.S.C. § 3323(b)
 5 CFR 930.201 et seq

15. Are you willing to travel? Yes No

Page 1
 OPM Form 1655
 September 2024

Figure 7. Spatial Text (S) representation of the same ALJ form. Whitespace characters preserve column alignment, allowing the model to associate values within the same table row (e.g., 01/15/2005, 06/15/2023, Administrativ, Departmen on the same line). Compare with the Plain Text in Fig. 6, where these values are serialized in separate column runs. This spatial structure accounts for the +3–18 pp accuracy gain reported in Sec. 4.3.